STRATEGIC PATENT BREADTH FOR DRASTIC INNOVATIONS

A Thesis Submitted to the College of Graduate Studies

and Research in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

in the Department of Agricultural Economics

University of Saskatchewan

Saskatoon

By

Amalia Yiannaka

Fall 2002

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirement for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Request for permission to copy or make use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Agricultural Economics,

University of Saskatchewan,

51 Campus Drive,

Saskatoon, Saskatchewan, S7N 5A8

CANADA

ABSTRACT

Patents are one of the most powerful forms of intellectual property protection and have been used for the last 500 years as a mechanism to stimulate the production and enable the dissemination of technological knowledge. Patents enable innovators to capture innovation rents through the granting of exclusive rights on their innovations for a pre-specified time period. The degree of appropriability of innovation rents enabled by a patent is determined, to a large extent, by the scope/breadth of the patent protection. In general, broad patent protection relaxes short-run competition in a market but it also increases the likelihood of a patent challenge and/or invalidation.

Despite the recognition that patent breadth has a negative effect on the likelihood of a patent remaining valid after it is granted, the standard assumption in the economics literature is that innovators attempt to maximize the returns from an innovation by claiming the broadest scope of patent protection possible. However, given that patent breadth is routinely challenged, the question arises as to whether the innovator, in an attempt to maximize the returns from the innovation, will choose the maximum patent breadth possible.

The purpose of this dissertation is to examine the optimal patent breadth strategy that an innovator should employ when faced with the possibility that the patent breadth claimed will be challenged. Specifically, the study argues that innovators of drastic product or drastic process innovations will not always choose the maximum patent breadth possible. Instead, in an effort to earn the maximum possible rents from the innovation, innovators will choose a patent breadth that is, in most cases, less than the maximum possible.

The dissertation uses a game theoretic approach to explicitly model the innovator's profit-maximizing patent breadth decision in two different situations – one in which the innovator has discovered a drastic product innovation and one in which the innovator has discovered a drastic process innovation. In both cases the innovator faces the possibility of infringement and/or a direct patent validity challenge.

Analytical results show that, contrary to what is traditionally assumed, it is not always optimal for the innovator to claim the broadest scope of patent protection. Instead, the strategic determination of optimal patent breadth involves a trade off. On the one hand, a broad patent protection increases the innovator's expected short-run returns by making it harder for competitors to enter his market without infringing the patent. On the other hand, a broad patent protection puts the viability of the patent at risk by increasing the probability that the patent will be infringed, that its validity will be challenged and that the courts will invalidate the patent or narrows its scope. Generally, this trade off implies that the innovator will choose a patent breadth that is less than the maximum possible. Under certain conditions, this less-than-maximum patent breadth will deter entry – however, deterring entry might not always be a profit-maximizing strategy for innovators.

The results of the game theoretic models are also used to develop the framework for an empirical model that could be used to examine the patenting behavior of innovators in different industries. While the empirical model is not estimated, the discussion in the dissertation indicates that estimation of the determinants of the patent breadth behavior is feasible and could be undertaken as a part of future research.

Understanding the factors that influence the determination of the optimal breadth of patent protection for the innovator and the effect of these factors on the profit-maximizing patent breadth may enable innovators to secure strong and viable patents. Holding strong and viable patents is particularly important for research intensive and innovative industries (i.e., biotechnology), where the ability to appropriate innovation rents is crucial for covering high research costs and guaranteeing further investment in the field.

ACKNOWLEDGEMENTS

The author wishes to thank those who contributed to the completion of this dissertation:

The College of Graduate Studies and Research for providing financial support.

Dr. Murray Fulton, my thesis advisor, for his continuous encouragement and thoughtful guidance and for his valuable advice on effective writing. I am also thankful to Dr. Fulton for his moral support especially during the completion of this dissertation.

Dr. Hartley Furtan, Dr. Richard Gray, Dr. William Kerr and Dr. Grant Isaac, members of my advisory committee, for the stimulating discussions during the early stages of this dissertation and for their very helpful comments and suggestions during its completion.

Dr. Brian Wright, for agreeing to become the external examiner for my dissertation, his time and very useful comments and suggestions.

Dr. Kien Tran, for his insightful suggestions on the empirical chapter of this dissertation.

Mike St. Louis, for his valuable help with computer related problems and his friendship.

All the faculty, staff and graduate students in the Department of Agricultural Economics for creating a stimulating and enjoyable working environment.

Father Agathagelos Kamvisis and Father Leonidas Iliakis, for their spiritual guidance and for giving me hope and strength during difficult times.

My brother Dinos, my sister Erie and my sister in-law Tala, who with their love, care, moral support and friendship, have encouraged and motivated me to do my best. I am also thankful to Dinos for our stimulating economic discussions and for helping me see the positive side of things.

Finally, my parents, Christos and Ritsa, for their unconditional love, encouragement and support and for always believing in me.

This dissertation is dedicated to my parents who were the first to show me the beauty and power of knowledge.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

Knowledge is expensive to produce, cheap to reproduce and difficult to profit from.

William D. Nordhaus

1.1 Problem Statement

Patents are an important mechanism for the generation and dissemination of technological knowledge. The development and growth of the agribusiness sector, and agricultural biotechnology in particular, has been enabled by the provision of patent protection. The emerging industrialization of agriculture and changing consumer preferences have made product differentiation and the development of cost reducing and/or quality enhancing processes important strategies for agribusiness firms (Drabenstott 1994, Boehlje 1996). The ability of innovating agribusiness firms to capture the returns of their innovations (e.g., new products and/or cost reducing/quality enhancing processes) is important for their economic performance and their incentive to invest in R&D.

Patents enable innovators to appropriate innovation rents through the granting of exclusive rights on their innovations. The limit of these exclusive rights is defined by two elements – patent length and patent breadth. Patent length is the time period during which the innovator has exclusive rights on the innovation and is predetermined by law. Patent breadth defines the technological territory claimed and protected by the patent and is explicitly chosen by the innovator. The innovator determines the breadth of the patent protection sought through the claims made in the patent application; the greater is the number of claims and the more general they are, the broader is the patent protection claimed.

The breadth of patent protection claimed is an important strategic variable for the innovator. Patent breadth determines, to a large extent, the innovator's ability to appropriate innovation rents. In general, the greater is the breadth of patent protection, the harder it is for competitors to enter into the patentee's market with non-infringing innovations and thus, the longer the innovator can enjoy the limited monopoly that the patent grants. At the same time, however, the greater is the breadth of patent protection claimed, the harder it is to secure a patent grant. In addition, even if a broad patent is granted, the greater is patent breadth, the greater is the probability that the validity of the patent will be challenged and that the courts will invalidate the patent or narrow its scope (Merges and Nelson 1990, Lerner 1994, Lanjouw and Schankerman 2001). A profit-maximizing innovator needs to consider both the marginal benefits and the marginal costs of an increase in patent breadth to determine strategically the optimal patent breadth claimed; the patent breadth that maximizes the innovator's ability to appropriate innovation rents.

Existing economic studies of the patenting behavior of the innovator/patent applicant have been limited to an analysis of the decision to either patent the innovation or to keep it a secret (Horstmann et al. 1985, Waterson 1990, Gallini 1992). Lerner

(1995) is the only study that empirically examines some other aspects of the innovator's patenting decision, namely, the decision to patent in certain subclasses. The innovator's role in the determination of the scope of protection covered by the patent (once the decision to patent has been made) has not been explicitly modeled in the literature. Instead, studies on patent policy and optimal patent breadth in particular explicitly or implicitly assume that the innovator tries to maximize innovation rents by claiming the broadest scope of patent protection possible (Gilbert and Shapiro 1990, Merges and Nelson 1990, Scotchmer 1991). The decision-making process of the patent applicant is thus assumed rather than derived in the existing studies.

1.2 Objectives of the Study

The principal objective of this study is to develop an understanding of the patenting behavior of an innovator who determines strategically the optimal breadth of patent protection claimed i.e., the patent breadth that maximizes the appropriability of innovation rents. The specific objectives are as follows:

- To examine/review the patent granting and patent challenge processes, as well as the economic literature on patent breadth, in order to provide evidence on and an understanding of the role of patent breadth as a strategic variable for the innovator.
- To explicitly model the determination of the optimal patent breadth for an innovator who has developed either a drastic product or a drastic process innovation and has decided to seek patent protection for this innovation.
- To identify the main determinants of the innovator's patent breadth decision and to determine their effect on the optimal patent breadth.

- To explore the effect of patent infringement and patent validity attacks on the innovator's patent breadth decision.
- To use the results of the theoretical models developed in this study to develop an empirical model that could be used, in a future study, to compare the observed patenting behavior in various industries to the strategic patenting behavior suggested by the theoretical models.

1.3 Methodology

The study develops two simple game theoretic models to examine the determination of the optimal patent breadth claimed when patent protection is sought for a drastic product and a drastic process innovation, respectively. In both models the patent breadth decision is modeled as a sequential game of complete and perfect information between an incumbent/patentee and a potential entrant. In both models the innovator acts strategically; he takes into consideration the effect that the patent breadth claimed has on the entrant and he realizes that the effort to safeguard his technological territory does not conclude with the granting of the patent. The innovator thus chooses the patent breadth that will induce the desired behavior from the entrant and he takes into consideration potential legal costs incurred when he enforces or defends his patent rights.

In the strategic patent breadth model for drastic product innovations the innovator moves first and decides on the breadth of patent protection claimed when he faces a positive probability that his patent may be infringed by the entrant. The potential entrant moves next and decides whether to enter the market and, if entry occurs, where to locate in the product space (i.e., to infringe or not to infringe the patent). In the last stage of the game the incumbent and the entrant choose their relative prices and compete in the market.

In the strategic patent breadth model for drastic process innovations the innovator moves first choosing the breadth of patent protection when he faces a positive probability that the validity of his patent will be directly challenged by a third party. The entrant moves next, after observing whether the patent has been challenged and the outcome of the challenge, if a challenge occurred. If the patent is challenged and revoked the entrant enters using the incumbent's process. If the patent is not challenged, or is challenged and upheld, the entrant decides how much to spend on research and development to generate her own non-infringing process. In the last stage of the game the players choose their respective quantities and compete in the market.

The study explores the innovator's patent breadth decision under different assumptions with respect to the environment in which the innovator operates. Thus, the patent breadth model for drastic product innovations assumes that the products produced by the two players are vertically differentiated, that the R&D process is deterministic and that patent infringement is a threat to the innovator. The patent breadth model for drastic process innovations, on the other hand, assumes that the processes produced by the two players are equally efficient and generate a homogeneous product, that the R&D process is stochastic, that a patent validity challenge is possible and that patent infringement is never a threat to the innovator.

Both models are solved using backwards induction. Backwards induction eliminates multiple Nash equilibria that do not represent credible threats that can exist in a sequential game. The equilibrium associated with the backward induction outcome is the only subgame perfect Nash equilibrium of the game (Fudenberg and Tirole 1991).

Finally, the theoretical findings of the two models, presented in the form of propositions, are used to develop an empirical model that could be used, in a future study, to examine the patenting behavior in various industries. The present study describes the selection of the patent sample that satisfies the theoretical assumptions and could thus be used in the empirical analysis, explores possible ways of approximating the variables of interest and identifies various data sources for these variables.

1.4 Organization of the Study

The remaining chapters of the dissertation are organized as follows. Chapter II describes the patent granting process, examines the possible ways a patent can be contested after it has been granted and demonstrates the economic importance of patent breadth for the innovator's ability to appropriate innovation rents. Chapter III reviews the patent literature and describes how patent breadth has been treated in the existing studies. Chapter IV develops a theoretical model to examine the patenting behavior of an innovator who, having invented a drastic product innovation and having decided to seek patent protection, determines the optimal patent breadth claimed. Chapter IV considers the possibility of infringement and entry deterrence. Chapter V develops a theoretical model to examine the patenting behavior of an innovator who, having developed a drastic process innovation and having decided to seek patent protection, determines the optimal patent breadth claimed. Chapter V considers the possibility of a direct patent validity challenge. Chapter VI uses the findings of the models developed in chapters IV and V to develop an empirical model that could be used to study the patenting behavior in various industries. Chapter VII summarizes the main findings and concludes the dissertation.

CHAPTER II

PATENT GRANTING, PATENT LITIGATION AND THE ECONOMIC IMPORTANCE OF PATENT BREADTH

2.1 Introduction

Patents are one of the most powerful forms of intellectual property protection and have been used for the last 500 years as a mechanism to stimulate the production of knowledge. They have played a crucial role in the development of research intensive and innovative industries, where the ability to appropriate the benefits from innovation is crucial for covering high research costs and guaranteeing further investment in the field.

Innovators seek patent protection as a means of safeguarding their technological territory from infringement and of relaxing competition in their market. A patent grants the innovator the right to exclude others from making, using, selling or importing the innovation without his/her permission. Therefore, it provides the innovator with monopoly rights on his/her innovation for a pre-specified time period. This time length represents the statutory life of the patent.

In return for the protection granted by the patent, the innovator has to disclose all information regarding the innovation. This information becomes publicly available and provides the building blocks for future research. The patent system thus provides incentives to innovate and is a mechanism for knowledge dissemination.

The granting of exclusive rights by the patent does not automatically imply protection and the ability to appropriate innovation rents, however. Instead, the main determinant of the protection granted by the patent and consequently of the appropriable rents is the patent scope or breadth.[1] Patent breadth determines the technological territory claimed and protected by the patent – the area in the technological space in which competitors cannot offer rival inventions without infringing the patent. The breadth of protection granted by a patent is case specific. Thus, unlike patent length, patent breadth cannot be pre-specified, pre-announced or standardized. Patent breadth is determined by the number and the nature (general/specific) of the claims that the patentee requests in the patent application and that are subsequently upheld by the Patent Office and/or the courts.

The innovator thus plays a crucial role in the patenting process and in the determination of the protection granted by the patent. The innovator's claims determine whether the patent will be granted and the breadth of patent protection. The breadth of the patent, in turn, determines the effective patent life. As will be discussed in more detail in the following sections, patent claims are often contested, either because the patent is infringed or because its validity is challenged. Court rulings on patent infringement and validity depend to a large extent on the claims that have been made. Given the above, the innovator's effort to safeguard his/her technological territory may not conclude with the granting of the patent.

The present chapter describes the workings of the patent system, examines the

---

[1] The terms width and height have also been used interchangeably with patent scope in the patent literature.

ways a patent can be contested after it is granted and demonstrates the economic importance of patent breadth for the innovator's ability to appropriate innovation rents. In specific, the present chapter consists of three sections. Section 2.2 focuses on the workings of the patent system. It begins with a short description of the evolution of the patent system and of the movements and treaties that have played a defining role in its establishment and expansion. A presentation of the principal objectives of the patent system and the requirements for patentability follows. Section 2.2 concludes with the description of the patent granting process and the Patent Office's and the innovator's role in the determination of the breadth of protection granted by the patent.

Section 2.3 examines the different ways a patent may be contested after it is granted. It describes the grounds on which the patent can be invalidated, what constitutes infringement, the different courses of action a patentee may take when his/her patent is infringed or the validity of his/her patent is challenged and the doctrines used by the courts when they rule on patent validity and infringement. Section 2.4 describes the importance of patent breadth for the level of innovation rents a patentee can appropriate with his/her patent. Finally, section 2.5 concludes the chapter.

2.2 Description of the Patent System

2.2.1 The Evolution of the Patent System

Economic historians trace the origin of intellectual property protection back to ancient Greece. Monopoly rights as a reward for inventive activity were first granted in Sybalis, an Italian town, as early as the third century BC. These monopoly rights were granted for a year to cooks of new recipes and were aimed at rewarding the inventor and promoting further inventive activity (Van Dijk 2000). The first organized patent system was instituted in the fifteenth century in Venice. In 1474, the Venetian Senate passed the first patent law, which granted ten-year monopoly privileges to inventors of new arts and machines (Cornish 1989). The practice of granting monopoly privileges to inventors of useful inventions was extended in many parts of Europe in the sixteenth and seventeenth centuries (Machlup and Penrose 1950).

The development of these early patent systems and their impact on the economy depended heavily on the nature and the purpose of the privileges that they granted. Some of the privileges that were granted offered protection from restrictive guild regulations and were aimed at reducing existing monopolies and at increasing competition (Machlup 1958). Other privileges protected the inventor against imitation and enabled the creation of monopolies. In other cases, the privileges granted had nothing to do with protecting and rewarding the innovator or encouraging the development of an industry. Instead, they served as a means of transferring profitable monopoly rights to favorites of the monarch. Patents of the latter nature were numerous in England in the 1560s and produced severe public criticism, which led in 1623 to the Statute of Monopolies (McKeough 1992).

The Statute of Monopolies ended the unrestricted granting of monopolies by forbidding the granting of exclusive rights to trade by the Crown, with the exception of patent monopolies to the first and true inventor of a new manufacture (Machlup 1958). Even though the Statute of Monopolies was not the first patent system, Machlup (1958, p. 3) refers to it as 'the Magna Carta of the rights of inventors' because it introduced the principle that only the first and true inventor could be granted a monopoly patent. Cornish (1989, p.73) on the other hand, argues that the objectives of the Statute were 'the encouragement of industry, employment and growth, rather than justice to the inventor for his effort.' Nevertheless, the law recognized the importance of inventive activity and that the granting of monopoly rights was a means to increase technological innovation.

The Statute of Monopolies formed the basis for the modern patent systems in most countries. After England, France and the United States were among the first countries to establish patent systems in 1791 and in 1793, respectively. Austria followed by introducing a patent system in 1810, Russia in 1812, Belgium and the Netherlands in 1817, Spain in 1820, Sweden in 1834, Portugal in 1837 and Saxonia in 1843 (Machlup and Penrose 1950).

The introduction and the expansion of the patent system did not occur without controversy, however. A strong antipatent movement rose during the period 1850-1875 in most European countries and was particularly powerful in the Netherlands, Switzerland, England and Germany. The antipatent advocates endorsed free trade and were opposed to anything that would hinder competition. Patent protection was linked to the establishment of monopolies and was thus anathematized by the free trade advocates.

Machlup and Penrose (1950) make reference to four basic arguments that were made by the patent system advocates in support of patent protection. The first argument was that every person should have the property right for his/her own ideas and the society should recognize and protect this right. The second argument held that it is only fair to give a just reward to a person for his/her contribution to the society. Since an inventor provided useful services (s)he should be 'appropriately' rewarded for them. According to the third argument, the patent system was an efficient and a cheap way of providing the necessary incentives for the generation and exploitation of innovations.

Finally, the fourth argument held that patents enable the disclosure of information and knowledge that would not have been publicly available in the absence of the patent system.

The opponents of the patent system offered convincing counter arguments as to why patent protection was undesirable. The main counter argument made with respect to the property rights argument was that patents frequently deny inventors the right to use their own inventions. This happens when many inventors come up with the same idea independently. In this case, only the first comer (the one who manages to patent the invention first) has the right to use it. Two counter arguments were made with respect to the 'just reward' argument. The first stated that no protection was necessary for the appropriation of innovation rents, while the second supported the idea of the reward but stated that prizes or bonuses were more efficient than patents. With respect to the third argument, opponents of the patent system claimed that the promise of a reward was not necessary to stimulate inventive minds. In addition, they claimed that the patent system was neither efficient nor cheap in providing research incentives. On the contrary, they claimed that there were substantial costs associated with the functioning of the system, legal costs of protecting and enforcing the patent right, as well as social costs from establishing temporary monopolies and costs from depriving others of the use of the invention. Finally, the main counter argument with respect to the 'secrecy argument' was that, given the substantial patenting costs and the uncertainty regarding the enforcement of the patent rights, inventors patent only when they cannot keep the invention a secret. Thus, patents were not necessary, as the information concerning the invention would have been disclosed anyway (Machlup and Penrose 1950).

The antipatent movement was initially successful and accomplished a reform of the patent law in England,[2] the extension of the delay of the introduction of a patent system in Switzerland and the abolition of the patent system in the Netherlands in 1869. But despite its initial strength, the antipatent movement did not survive. The severe depression of 1873 weakened the free trade argument, which in combination with the organized counterattack by the advocates of the patent system, led to the disappearance of the antipatent movement. As a result of the success of the propatent movement, England withdrew the patent reform of 1874 and Germany adopted a uniform patent law in all its territories in 1877. Switzerland instituted its first patent system in 1882, mainly to avoid the stigma of piracy attached to the nations that refused to introduce a patent system, and the Netherlands re-established its patent system in 1912 (Machlup 1958).

The establishment of patent rights in the rest of the world was enabled by the Paris Convention for the Protection of Industrial Property of 1883. The treaty, which today has 140 signatories, provided that each member country guarantees to the citizens of other countries the same treatment that it provides to its own citizens.[3] If a country did not have a patent system, its nationals could not seek patent protection in the countries that granted protection (Van Dijk 2000). In addition, the treaty established a system of patent priority. Under this system, an application in one of the member countries allowed for a period of twelve months during which another application could be made in any other member country that would bear the same priority date as the first application (Cornish 1989).

Further expansion and improvement of the patent system was enabled by the

---

[2] The patent law reform of 1874 instituted compulsory licensing of all patents, required stricter examination of patent applications, reduced patent life to seven years and nullified patents that were not used for two years after grant (Machlup and Penrose 1950).

[3] The Convention was revised in Brussels in 1900, in Washington in 1911, in Hague in 1925, in London in 1934, in Lisbon in 1958 and in Stockholm in 1967 (Cornish 1989).

development of the World Intellectual Property Organization (WIPO), which today has 175 member states and administers 21 treaties on intellectual property protection.[4] WIPO's objectives regarding patent protection are the harmonization of national patent laws and procedures, the provision of international patent applications, the provision of legal and technical assistance to countries, the exchange of intellectual property information and the resolution of disputes among countries (WIPO 2001a). Two of the treaties administered by WIPO, the Patent Cooperation Treaty (PCT) and the Patent Law Treaty (PLT), work towards the achievement of the above objectives.

The PCT that came into force in 1978 and today has 108 signatories established a system of international patent cooperation. This system allows an applicant to initiate patent protection procedures in several countries by filing one international application. The PCT system has been very successful; the number of international patent applications has grown from 2,600 in 1979 to 74,000 in 1999 (WIPO 2001b). The international application has the same effect as the national filing in each member country in which protection is sought.

The PCT allows for an international search and it optionally provides an international preliminary examination. It provides the inventor with a twenty-month period after the international filing (or a thirty month period if preliminary examination is requested) to file a national application, a period that can be useful in assessing the commercial possibilities of the invention in the countries of interest. The PCT is only a patent filing procedure, which does not grant an international patent. The granting of a

---

[4] WIPO has its roots in the Paris and the Berne Conventions. In 1893 the international bureaus of the above Conventions were united to form the United International Bureaus for the Protection of Intellectual Property (BIRPI). In 1970 BIRPI underwent structural and administrative reforms and became WIPO (WIPO 2001a).

patent remains the responsibility of the national Patent Office of each country in which protection is sought.

The PLT was established in June 2000 and it has 48 signatories. The treaty was designed to harmonize national patent laws and procedures regarding the application, the granting and the maintenance of patents. The PLT incorporates the requirements of the PCT international applications in national and regional patent laws. The harmonization of patent procedures will contribute to easier access to worldwide patent protection and to a reduction in patenting costs. It is also expected to contribute to the efficient operation and to a reduction in the administrative costs of the Patent Offices of the member states (WIPO 2001c).

2.2.2 The Objectives and the Requirements of the Patent System

The main objectives of the modern patent system are to foster technical innovation, to diffuse new technological knowledge by achieving early innovation disclosure and to avoid R&D duplication (EPO 2000a). It is commonly believed and often claimed by economists that rewarding inventive activity is another principal objective of the patent system (Gilbert and Shapiro 1990). Cornish (1989), however, regards the idea that the patent system is 'an instrument of justice to the inventor' as incorrect. Instead, he considers that rewarding the innovator is 'an incidental consequence of modern patent systems,' not a principal objective (Cornish 1989).

Nevertheless, rewarding the innovator is an important condition for stimulating the production of technical knowledge and achieving early information disclosure. Thus, through the provision of exclusive rights over the invention which generate a limited monopoly, the patent system provides incentives for the undertaking of research and development. The patent system transforms technical knowledge from a public good (i.e., nonrival and nonexcludable) to a toll good (i.e., nonrival and excludable), thus allowing the 'first and true' inventor to appropriate innovation rents that would have been difficult or even impossible to appropriate in the absence of the system (Miller and Davis 1990). In return for the protection granted by the patent, the innovator has to disclose all information relative to the innovation. A patent can be thus viewed as a social contract between an inventor who agrees to disclose all information regarding the invention and the government who, in return for the information supplied, agrees to offer protection from competition (Machlup 1950).

The patent system is by no means an optimal mechanism of knowledge generation however, as it leads to a price of knowledge higher than its marginal cost (of zero) for as long as the patent is in effect. Therefore, even when information disclosure is perfect, there is a deadweight welfare loss generated by the limited monopoly that the patent establishes. Nor is the patent system an optimal mechanism of safeguarding intellectual property. Mansfield et al. (1981) have found that sixty percent of the patented innovations that they studied were imitated within four years of their introduction. The innovator must thus often protect or defend his/her patent right after the patent grant.

Despite its inefficiencies, the patent system has managed to survive through the years as a strong feasible alternative to other forms of protection. In some industries (i.e., biotechnology) it is a preferable alternative to the no intellectual property protection outcome. As Lerner (1995) points out, trade secrecy is not a good alternative to patent protection for the biotechnology industry due to the strong academic roots of the industry and to the high mobility of its human capital.

The rights granted by a patent refer to a specific time period which represents the statutory life of the patent. The statutory life of the patent is twenty years in most countries. The patent rights extend only through the territory of the country (or the region in the case of patents granted by the European Patent Office (EPO)) that granted the patent and they have no effect in another country (or region). A patent is granted to any person who, in the language of the statute, 'invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement subject to the conditions and requirements of the law' (USPTO 1999, EPO 2000a, CPO 1998g).

A process is a method of generating an existing or a new product; it is a means to an end. Process innovations are cost reducing and/or increase the efficiency of producing a given product. The other three categories, machine, manufacture and composition of matter, are the ends of a process and they are all classified as products. Product innovations are new products and/or improvements of existing products. Essentially, there are only two categories of patentable subject matter, processes and products (Miller and Davis 1990). If a process is patentable, it doesn't mean that the product that it generates needs also to be patentable, and vice versa.

Non-patentable subject matter includes abstract ideas (even though the application of an idea may be patented), mental steps, laws of nature, principles and fundamental truths, printed media matter, business methods and naturally occurring products. A combination of naturally occurring products may be patented, as long as it results in a new non-naturally occurring product.

The fundamental patentability requirements according to the patent law are novelty, utility (for most patents) and nonobviousness. Patents for which utility is a basic requirement are known as utility patents. For plant and design patents, the requirement of utility is substituted by the requirement of distinctiveness and ornamentality, respectively (USPTO 1999).

The requirement of novelty establishes that an invention cannot be patented if it is not new, that is, if it forms part of the prior art (CPO 1998f). An invention is not considered to be novel if it is anticipated. Anticipation of the invention occurs when it was in public use or on sale or when it was previously patented or described in printed publications anywhere in the world prior to the claim date. The claim date of the patent application is the filing date of the application. Prior publication will bar issuance only when all features of the invention have been disclosed in a single prior publication. In other words, references may not be combined to show lack of novelty, since this consist an improper 'mosaic' of references (CPO 1998c).

The novelty requirement is also violated when certain events occur that raise statutory bars. Statutory bars are raised when the invention was in the public use or on sale in the country where the application was made or when it was patented or described in a printed publication anywhere in the world for more than one year before the application for a patent was made. The inventor is given a grace period of a year after the occurrence of an event that raises the statutory bar during which (s)he may apply for the patent. If (s)he fails to do so, and the grace period is over, the inventor loses his/her application priority and the right to a patent (Miller and Davis 1990).

It is immaterial whether it is the innovator or another party who causes an event that raises the statutory bar. The role of the statutory bar is to induce inventors to apply for patent protection early so that they can protect their inventions from piracy and so that early information disclosure can be achieved. Miller and Davis (1990) state that a patent will not be granted if the benefits of the invention have been disclosed to the public, even when the secrets of the invention have not been disclosed. The patent system does not favor the granting of protection to an invention whose benefits have already been made public.

Utility is another requirement for patentability. Even if an invention is found to be novel it will not be patented if it is not found to be useful. Utility related to inventions usually means industrial value. The utility of the invention must be clearly demonstrated in the patent specification. The purpose of the utility requirement is to avoid the patenting of an illegal, immoral or detrimental invention (USPTO 1999, EPO 2000a). In addition, the requirement of the demonstration of useful applications of an invention helps to define the scope of the patent protection.

Similar to the novelty requirement, the nonobviousness requirement is decided based on the prior art. There is a significant difference between the novelty and the nonobviousness requirement, however. Novelty considers the entire prior art with respect to substantial identity, while nonobviousness considers only the prior art that is relevant to the invention with respect to what would have been the next obvious step (Miller and Davis 1990). Thus, an invention such as a combination of chemical compounds that have never being combined in this particular form before may be novel but it may also be obvious. An invention is considered obvious when a person skilled in the relevant art could have invented it with relative ease had he tried to do so (EPO 2000a).

The process of determining nonobviousness consists of three steps. The first step consists of determining the prior art that is relevant to the invention. The second step consists of the identification of the differences between the prior art and the invention.

This second step will determine whether an inventive step has occurred. Finally, the third step requires the determination of the level of skill of an ordinary practitioner in the relevant art. Miller and Davis (1990) include a fourth step to the above process; the secondary considerations. The existence of secondary considerations is an indication of the nonobviousness of the invention. According to Miller and Davis (1990, p.87) 'the secondary considerations include commercial success, long felt but unsolved needs, the failure of others, the fact that experts express disbelief at a discovery's announcement, the very existence of well established prior art and later improvements upon that discovery.'

2.2.3 The Patent Granting Process

Having invented what the innovator believes to be a novel, nonobvious and useful invention, the innovator may apply for a patent. Given that every country has its own patent laws, a patent application should meet the requirements of the laws of the particular country in which the application is made. Patent granting procedures that are common to the Canadian Patent Office (CPO), the EPO and the United States Patent and Trademark Office (USPTO) are presented in what follows. Differences in the practices of these Patent Offices will be noted.

2.2.3.1 The Patent Application

A patent application consists of the following basic parts: an application transmittal form; the title of invention; cross references to related applications (if any); background of the invention; brief summary of the invention; brief description of the several views of the drawing (if any); detailed description of the invention; claim or claims; abstract of the disclosure; drawings (if any); executed oath or declaration and sequence listing (if any) (USPTO 1999, EPO 2000a). The most important part in the patent application is the patent specification. The patent specification consists of the patent description and the patent claims. The description discloses the invention while the claims mark out the scope of the monopoly rights (Cornish 1989).

- *The Patent Description*

The patent description illustrates the problem faced by the innovator and the steps followed towards its solution. The description consists of the title of the invention, a description of the technical field to which the invention pertains and a description of the relevant prior art. The description of the relevant prior art, including citations of all previous patents related to the invention, is of great importance as it helps to distinguish the invention from prior art and provides a better understanding of the invention (Lanjouw and Schankerman 2001). A description also includes a complete explanation of all embodiments of the invention and a description of at least one way for carrying out the invention in the form of examples, drawings and figures, where appropriate (CPO 1998b).

The description of the invention and of the manner and process of making and using it must be clear, concise and exact. It should disclose the invention in its best mode, so as to enable any person skilled in the relevant art to reproduce the invention without further experimentation. The best mode requirement is an important, albeit subjective, requirement since it means that the description need not provide the best mode that is theoretically available, but rather the best mode of which the inventor is aware (Miller and Davis 1990).[5] The description should fulfill an important statutory requirement, that of providing an enabling disclosure. If the description is broad, obscure or ambiguous and further experimentation is required for the reproduction of the invention by those skilled in the relevant art, then the disclosure is inadequate, nonenabling and will be found to be invalid. Finally, the description should describe all subject matter that the applicant intends to claim as his/her invention (CPO 1998b).

- *The Patent Claims*

The patent claims conclude the patent specification and are its most important element. Not only do they define the subject matter that the innovator regards as the invention, they also determine the technological territory for which protection is sought. According to Miller and Davis (1990), patent claims define the perimeter of the invention. The claims must be drafted very carefully, since the wording of claims determines to a large extent whether a patent will be granted and the scope/breadth of patent protection (USPTO 2001a, EPO 2000a, CPO 1998c). The claims continue to be important after the patent has been granted, since courts rule on questions of validity and infringement on the basis of claims (USPTO 1999, Merges and Nelson 1990).

Claims can be independent and/or dependent. The independent claims define the technological territory claimed in a broad form, while the dependent claims are more restricted and describe detailed features of the invention. According to the definition of

---

[5] The importance of the best mode requirement is reflected in the following court ruling in the case of Scripps Clinic & Research Found v. Genentech (666 F. Supp. 1379 3 U.S.P.Q. 1481 (N.D. Cal. 1987). Scripps Clinic held a product patent on a purified blood clotting protein and accused Genentech for infringement for producing the same protein with recombinant DNA techniques. Even though at a first court ruling the Genentech product was found to infringe the patent at a later court ruling the accused infringer succeeded in proving that the original patent was invalid on the grounds that the best mode of operation was not revealed by the patent (Merges and Nelson 1990).

the EPO (2000a), a claim is dependent if it includes all the features of any other claim. Dependent claims are drafted with reference back to and further limiting other independent claim or claims in the same application. Dependent claims should be grouped together (USPTO 2001a). A patent specification must contain one or more independent claims provided that they are sufficiently different from each other (USPTO 2001a).[6] Dependent claims are construed so that they become progressively narrower. In other words, a succession of claims covers increasingly specific areas (Cornish 1989).

The smaller the number of independent claims and the greater the number of the elements within a claim (the more specific is the claim), the narrower is the patent protection. Miller and Davis (1990, p. 113) state that 'in a patent claim more means less.' The view that more detailed claims imply less protection is known to patent law practitioners as the *more is less* principle (Matutes et al. 1996). The justification for this belief is that the more general are the claims upheld by the Patent Office, the greater is the probability that another invention that incorporates the property described in the claim will violate the patent.

Patent claims should be drafted so as to achieve two conflicting goals; they should demonstrate patentability (i.e., how the invention meets the novelty, utility and nonobviousness requirements) and they should define what would constitute infringement (Miller and Davis 1990). To demonstrate patentability, the claims should be clear, concise and they should be supported by the description (EPO 2000a). They can be as narrow as the applicant wishes, but they should not be broader than the

---

[6] Plant patent specifications should contain only one claim since a plant patent is granted on the entire plant (USPTO 1999).

invention as described or supported by the description (CPO 1998c).[7] The broader are the claims, the harder it is to demonstrate that the invention is different from the prior art (i.e., that it is to show novelty and nonobviousness) and that the description is enabling (Merges and Nelson 1990). At the same time, to offer protection from infringement the claims should not be so narrow as to give up what should legitimately be the property of the inventor (Miller and Davis 1990).

It is of great importance that the claims are supported by the description as 'the specification is the dictionary to which the courts refer when the language of the claim is questioned' (Miller and Davis 1990, p. 112). The requirement that the claims must be supported by the description does not mean that the description must indicate precisely how to make every product or process that would fall within its claims, however. Merges and Nelson (1990, p. 846) state that '[d]isclosure of an inventive concept or principle, whose precise contours are defined by the claims, is enough.' In other words, at the time of the invention and of the application for a patent, the inventor cannot possibly be aware of all the embodiments of his/her invention. Thus, (s)he should be entitled to embodiments that fall within his/her claims but are not specifically described in the specification.

The Patent Office may object to claims due to lack of support or nonenablement only if it can show that a person skilled in the relevant art would be unable to reproduce

---

7 The EPO (2000a) in its "Guidelines for Examination in the EPO" gives examples of claims not justified by the description. The EPO states that when this happens, the patent may be objected to either because of lack of support or insufficient disclosure. An example given by the EPO (2000a, part C, Chapter III, p. 39) is the following: 'A claim relates to a process for treating all kinds of "plant seedlings" by subjecting them to a controlled cold shock so as to produce specified results, whereas the description discloses the process applied to one kind of plant only. Since it is well known that plants vary widely in their properties, there are well-founded reasons for believing that the process is not applicable to all plant seedlings. Unless the applicant can provide convincing evidence that the process is nevertheless generally applicable, he must restrict his claim to the particular kind of plant referred to in the description. A mere assertion that the process is applicable to all plant seedlings is not sufficient.'

the invention based on the information disclosed in the application without undue experimentation. The more pioneering/drastic is an innovation, the harder it is for the Patent Office to object to broad claims, since the harder it is to find something in the prior art to demonstrate that embodiments of the claimed invention would be impossible to make without undue experimentation.[8] Thus, for drastic innovations the burden falls on the Patent Office who must disprove enablement (Merges and Nelson 1990). In addition, according to the EPO (2000a, part C, chapter III, p. 38) 'an invention that opens up a whole new field is entitled to more generality in the claims than an invention that is concerned with advances in a known field of technology.' Thus, claims to drastic inventions are often allowed to cover areas beyond the area that is examined and disclosed by the inventor. The narrowing of the claims of pioneering/drastic inventions is typically left to the courts (Nelson and Merges 1990).

The number and the nature of claims allowed by the Patent Office are case specific and depend on the prior art. The Patent Office has considerable room for discretion when it decides which claims should be admitted, narrowed or rejected (USPTO 1999, EPO 2000a, CPO 1998e).[9] The patent examination and granting process

---

[8] Drastic innovations are processes that reduce cost or products that generate new demand or meet demand not previously met so that the innovator can exert monopoly power within the market. Merges and Nelson (1990, p. 854) give a US Supreme Court definition of a patent covering a pioneer invention. According to this definition, a patent covering a pioneer invention is 'a patent covering a function never before performed, a wholly novel device, or one of such novelty and importance as to mark a distinct step in the progress of the art'.

[9] An example demonstrating that there is room for discretion when examining patent applications is given by the transgenic mouse patent (known as the Harvard mouse patent). The patent was granted in 1988 by the USPTO (U.S. Patent No. 4,736,866) to doctors Phillip Leder and Timothy Stewart of the Harvard Medical School. The innovators claimed protection not only for the process and the transgenic mice variety that they produced but also for every non-human mammal that would be produced by their method. The USPTO accepted the inventor's claims that their process could be used to engineer any non-human mammal. The EPO on the other hand, rejected the claims that claimed subject matter beyond mice and rodents on the basis of being nonenabling. The EPO stated that '[a]nimals which have been used in the prior art are mainly mice and no instructions are to be found in the specification as to how success could be achieved with other mammalian animals' (Merges and Nelson 1990, p. 848).

followed by the Patent Office when it decides on the nature (general/specific) and the number of claims finally allowed, if any, is presented in the next subsection.

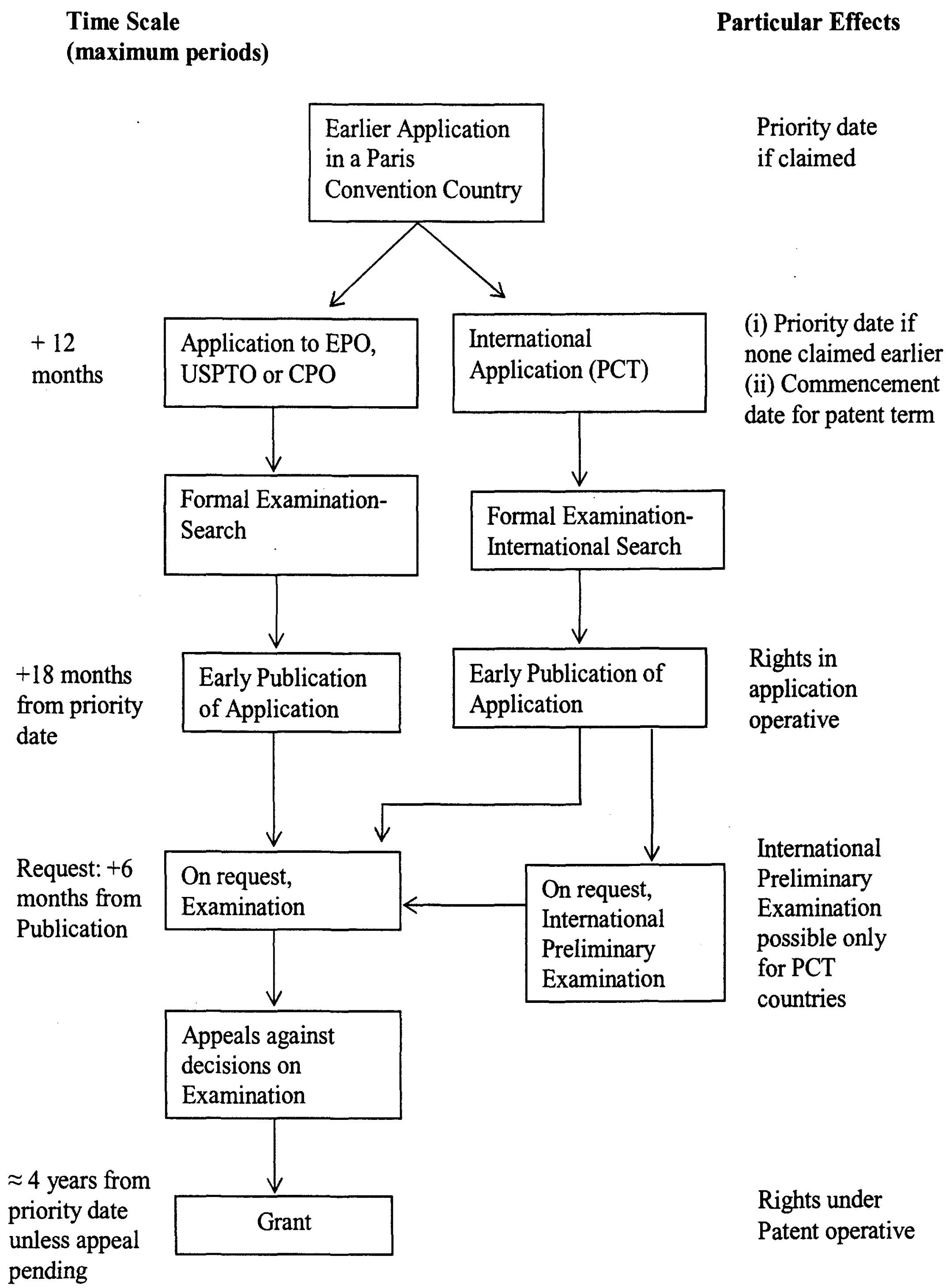
2.2.3.2 The Patent Examination and Granting Process

The patent examination and granting process is depicted in Figure 2.1. The formal examination of the application by the Patent Office takes place as follows. The applicant submits his application with a request for a search and a preliminary examination (optional). Upon submission, a classification examiner classifies the application in sections, classes and subclasses, groups and subgroups based on the national and the international patent classification (IPC) system, if they are different.[10] The proper classification of each application is of great importance, since this classification determines the examining group that will review the application and the area of prior art that the examiner will search to determine the patentability of the invention (especially novelty and nonobviousness). If the application is made to the EPO or it is an international patent application, then the applicant must designate the countries in which protection is sought.[11]

---

[10] The CPO and the EPO use the IPC system. The IPC system is industry and profession oriented. The USPTO uses both the IPC and its own classification. The USPTO classification is structure and function oriented (Lerner 1994). The CPO (1998d) in its 'Manual of Patent Office Practice' gives examples of the IPC classification mode. According to the examples an IPC A61K 31/025 is read as follows: "A" is the Section,"61" is the Class, "K" is the Subclass, "31" is the Main Group, and "025" is the Subgroup.

[11] The designated countries may be withdrawn later on but they cannot be added to (Cornish 1989).



Source: Adopted from Cornish (1989).

Figure 2.1 Patent Examination and Grant

After a search has been conducted, an early publication of the application takes place for the benefit of third parties at 18 months from priority date (CPO 1998a, EPO 2000a, USPTO 2001b). From this point onward, the Patent Office file is open to inspection. During this stage a third party may communicate with the Patent Office to raise issues regarding the patentability of the invention at issue or concerning prior art of which the Patent Office might be unaware. A third party, however, does not have the right to present a case in opposition to the grant at this stage. In addition, no information regarding the action taken by the Patent Office will be given to a third party. Since, during this stage of the process, a third party cannot make a case on how the provided information should be used by the Patent Office, this strategy of interfering with the granting of a patent is usually not attractive to competitors (Cornish 1989). Instead, competitors prefer to oppose the patent after it has been granted, a process described in subsection 2.3.2.

After receiving the search report the applicant may amend his application on his/her own initiative explaining the nature of the amendments and their purpose. The applicant is not allowed to add to the original description, but (s)he is allowed to amend his/her claims. A written request for examination must be submitted to the Patent Office for the examination of the application to commence. The request may be submitted along with the application or after the applicant receives the search report from the Patent Office. If the latter option is chosen, the written request for examination must be submitted within a pre-specified time period (usually within a six month period after the search report has been published).

The EPO (2000a) recommends that the patent examiners conduct a second search ("topping-up" search) after studying the claims to determine the existence of conflicting applications. If, during examination, the patent examiner finds that there are other patents currently pending or granted within the previous year that make the same (or similar) claims, (s)he should initiate an interference proceeding (Miller and Davis 1990). The interference proceeding will determine which of the applications should be given priority. Priority for obtaining a patent is determined in most countries by 'the first to apply' rule. In the United States, the party that is the first to invent the innovation that is given priority.[12]

When all necessary searches have been completed the invention is examined in light of the search reports to determine whether it meets the statutory patentability requirements and whether the description and the claims are in line with the patent rules. The Patent Office may, after examination, accept the application as filed, object to it and make a request for amendments or it may find that the invention does not meet the patentability requirements and may reject the application. In the case that the patent examiner has no objections, a notice of allowance is issued to the applicant who then has to pay a final fee to obtain his/her patent.

When the examiner has objections with respect to the compliance of the application to the rules or (s)he finds the subject matter nonpatentable, (s)he issues a report requesting the applicant to amend his/her application or explains the reasons for rejection. The examiner has to point out which claims are objected to or rejected. The examiner must also specify the time limits (usually six months) within which the amendments should take place (Cornish 1989). The applicant must then request

[12] A proof of priority in the conception of an invention may be established by submitting a disclosure document to the USPTO. The disclosure document describes the invention and explains the way it is produced and used. The disclosure document is usually submitted prior to the patent application. It will be retained in confidence by the USPTO for a period of two years and it will be used to prove priority of conception of the invention.

reconsideration in writing. The applicant is expected to reply by a written statement to every ground of objection and rejection. The applicant must explain the manner in which the amendment overcomes each of the objections and/or rejections made by the examiner or, when amendments are not made, to provide arguments as to why the application in its original form does comply (Miller and Davis 1990, CPO 1998e). The amendments may not introduce new subject matter. At this stage amendments on the initiative of the patent applicant cannot be made without permission from the Patent Office.

Amended applications are subject to further examination and the applicant will be informed as to whether the claims are rejected, objected to, or allowed. If the examiner has no further objections a notice of allowance will be issued. In some cases an impasse is reached between the applicant and the examiner during the process of amending the application. In this case, the application is rejected in a final action, which terminates the prosecution of the application (USPTO 1999, CPO 1998h, EPO 2000a). The final action report must clearly indicate which claims are allowable (if any) and which are still objectionable. The applicant may then appeal the final decision to a Board of Patent Appeals. When the Board's decision is still adverse, the applicant may further appeal to the courts.

Statistics from the USPTO and the EPO show that patents are finally granted to two-thirds of the applications (USPTO 1999, Van Dijk 2000). The EPO (2000b) states that it takes on average four years for a European patent to be granted. After the patent has been issued it is granted protection for the maximum of twenty years from the filing date as long as the required renewal fee is paid. The fees charged are a function of the length of protection; the closer the patent gets to its expiry date the higher is the fee. The patent will lapse as a result of non-payment of the renewal fees within a specified time period (usually six months).

The granting of a patent does not guarantee that the patent will stay active for its statutory life or that its scope/breadth will stay unchanged. Jewkes et al. (1960) state that nine out of ten patents do not remain active for the period that the patent is legally protected. Data on patent renewals shows that more than fifty percent of patents are voluntarily cancelled by non-payment after ten years of their application (Scotchmer 1999, Cornelli and Schankerman 1999). Voluntary cancellation of the patent primarily takes place when the patent ceases being profitable, as a superior technology has rendered the patented one obsolete.

The effective patent life may also be shorter than the statutory patent life as a result of non-voluntary cancellation of the patent. This happens when a patent is found to be invalid and it is revoked in the course of a direct patent validity challenge and/or during an infringement trial. Patent challenges and litigation also affect the breadth of protection that the patent grants – this breadth may be narrowed even when a patent is not invalidated. Section 2.3 that follows examines how a patent may be contested after it has been granted.

2.3 Patent Validity Challenge and Patent Infringement

Innovators seek patent protection as a means of appropriating innovation rents. The inventors' effort to safeguard their technological territory does not usually conclude with the granting of the patent, however. Jewkes et al. (1960) state that the critics of the patent system view the patent right solely as a right that must be defended in the courts, while Cornish (1989) claims that the precise scope of a valuable invention is likely to be the subject of constant questioning.

A patent may be contested in different ways after it has been granted. A patent is contested either when the patented subject matter is infringed or when its validity is directly challenged (Cornish 1989). In the case of patent infringement, the patentee has the option to try to enforce his/her patent rights by filing an infringement suit.[13] In the course of the infringement trial the validity of the patent may be indirectly challenged; in this case the challenge takes the form of a defensive counterclaim (see subsection 2.3.1). The validity challenge may also take the form of an offensive action in which an opposition is launched in either the Patent Office that granted the patent or in the courts (i.e., direct validity challenge). Contesting a patent – both directly and indirectly – could lead to the narrowing of the breadth of its protection or even to the elimination of protection as a result of patent invalidation. The patent is found to be invalid when it is not found to be novel, useful, nonobvious or enabling.

Patent oppositions and litigation have escalated during the last decades (Barton 2000, Hofer and Edmundson 1999, Lanjouw and Schankerman 2001). One of the main reasons behind the increase in patent oppositions and litigation is considered to be the quality and amount of resources available to the Patent Offices. Thus, while patent applications have increased substantially, the resources of the Patent Offices are limited.[14] Barton (2000) claims that a patent examiner spends on average only twenty five to thirty hours examining an application. The limited time spend on examining a

[13] The patentee may choose to not take this option if (s)he fears that his/her specification is 'weak' and (s)he might loose in court.

[14] Voss (1999) claims that the number of applications in the USPTO increases by more than 8% annually, mainly due to software and biotechnology applications. Phillips and Dierker (2000) state that biotechnology patent applications in Canada have increased by 12.9% annually during the last decade.

patent adversely affects the ability of the examiner to conduct effective searches and to evaluate the patent claims. In addition, the *ex parte* nature of the patent prosecution process leads to the Patent Office often granting broad claims that cannot survive a validity challenge. Lerner (1994) states that the USPTO often awards patents that appear to overlap, leading to disputes that are either resolved through negotiations or litigation.

The introduction of intellectual property protection to new technology areas like computer software and biotechnology has also contributed to the increase in patent oppositions and litigation. The lack of prior art (given that these are new areas) where the patent examiner could seek support to object to erroneous and broad claims and the complexity of the claims in these areas combined with the limited resources at the Patent Offices, intensify the problem. In addition, there is little legal precedence established in new industries that could help disputants form realistic expectations with respect to the outcome of the challenge and/or trial which leads to more oppositions and/or litigation (Lanjouw and Schankerman 2001).

Section 2.2 established the importance of claims in determining the patentability of the invention and in defining the scope/breadth of the protection granted by the patent. The present section demonstrates that claims continue to be important after the granting of the patent since questions of infringement and validity are judged, to a large measure, on the basis of claims.

The breadth of patent claims is basically determined by three doctrines, the enablement doctrine, the doctrine of equivalents and the doctrine of reverse equivalents (also known as the doctrine of literal overlap) (Merges and Nelson 1990). The enablement doctrine, which has been examined in Section 2.2, is of importance when the validity of the patent is under scrutiny. The doctrines of equivalents and reverse equivalents are applied when infringement is examined. The application of the doctrine of equivalents broadens the scope covered by the claims while the application of the doctrine of reverse equivalents narrows it. The subsections that follow outline how the above doctrines are used when questions of patent infringement and validity are examined.

2.3.1 Patent Infringement

Direct infringement of a patent occurs when someone produces, uses or sells the patented product or process without the permission of the patent owner. Anyone who encourages an infringement of a patent is liable for indirect infringement, while anyone who supplies or sells an item whose only or principal use is in connection with a patent is liable for contributory infringement (Miller and Davis 1990).

A patentee that believes that his/her patent has been infringed may protect his/her rights by suing or threatening to sue the alleged infringer. The threat of suing can be very effective in some cases as it may cause the alleged infringer to terminate any infringing action. To gain time, the patentee may ask the courts to grant a preliminary injunction (also known as an interlocutory injunction). A preliminary injunction orders the accused infringer to discontinue the allegedly infringing action until trial. Preliminary injunctions are not very common, as it is very hard to prove before trial that the patent is valid and it has been infringed. For a preliminary injunction to be granted the patentee must also prove that in the case that (s)he losses at trial, (s)he will compensate the alleged infringer for any damages that the latter may have suffered from the injunction.

On the other hand, an alleged infringer who believes that (s)he has been wrongfully accused or maliciously attacked by a patentee has the right to protect himself/herself by establishing 'freedom from liability' (Cornish 1989). In the case that the alleged infringer has been threatened with a lawsuit, (s)he can ask the courts to declare that the threats are unjustifiable, for an injunction against the continuance of the threats and for damages for any losses that the threats have caused. Threats to sue for infringement can cause considerable losses to the accused infringer due to loss of customers to the threatener (Merges and Nelson 1990).

Patent infringement proceedings can be lengthy, uncertain and costly for all the parties involved. The American Intellectual Property Law Association (AIPLA) in its 1999 Economic Report estimates average litigation costs to be 1.5 million dollars per side (Barton 2000). The party who fails to prove its case at trial will be asked to pay for any damages caused to the other party because of its actions and in some exceptional cases to even pay the other party's attorneys' fees (Miller and Davis 1990).[15]

If the patentee wins at trial, the infringer will pay damages that may either represent lost royalties/lost profits caused by the infringing action or profits realized by the infringer. When the doctrine of lost profits/lost royalties is applied, the aim is to compensate the patent holder by restoring his/her 'but for' the infringement position so as to maintain his/her incentives to invest in R&D (Schankerman and Scotchmer 2001). When damages are awarded based on the doctrine of unjust enrichment, the idea is to

[15] Attorneys' fees may be granted when it is proven that the infringer has willfully infringed the patent (he believed it to be both valid and infringed by his/her actions) or when the plaintiff has brought an infringement suit even though (s)he did not believe his/her patent to be valid and infringed (Miller and Davis 1990).

enable the patentee to recover profits realized by the infringer's wrongdoing.[16]

If the accused infringer wins at trial the patentee will pay damages for lost profits caused by the infringement suit or for any harm caused by the threats to sue. Not all lawsuits go to trial, however, since trial is an expensive way to resolve patent disputes. The parties involved in a lawsuit may rather try to resolve their differences in a pretrial settlement. Pretrial settlements can also be very costly. Hofer and Edmundson (1999) claim that in some cases settlement costs have exceeded one hundred million dollars.

The validity of the patent often becomes an issue during an infringement trial. At trial the patentee will try to prove that his/her patent is valid and that it has been infringed. The alleged infringer on the other hand will try to first prove that the patent is invalid and second that the alleged infringing device does not infringe the patent (Merges and Nelson 1990). The accusation that the patent is invalid may be just a form of defense or it can take the form of a counterclaim demanding the revocation of the patent (Cornish 1989).

Establishing infringement can be a formidable task. Whether infringement will be found depends on the way that the patent claims are interpreted. The test used by the courts to determine whether a product or process infringes the patent is first to examine the allegedly infringing product or process in the light of the patent claims to determine whether the claims describe the product/process. If the claims describe the product/process, meaning that the product/process falls exactly within the boundaries of patent claims, then there is literal infringement of the patent (Merges and Nelson 1990).

[16] An example of the award of damages based on the doctrine of unjust enrichment is given by the case of Polaroid vs. Eastman Kodak. In this infringement lawsuit, Polaroid won the right to put Kodak out of the instant camera business (with a permanent injunction) and was awarded damages of over 873 million dollars (Hofer and Edmundson 1999).

If literal infringement is not found, then the courts examine whether infringement occurs under the doctrine of equivalents.

The doctrine of equivalents asserts that 'if two devices perform the same functions in *substantially* the same way and accomplish *substantially* the same result, they are the same, even when they differ in shape, name and form' (Merges and Nelson 1990, p. 853).[17] The application of the doctrine of equivalents allows infringement to be found even when the accused device does not fall within the patent claims. In other words, it requires that the accused device is substantially equivalent and not absolutely identical to the patented one for infringement to be found (Miller and Davis 1990). The doctrine of equivalents narrows the possibilities of infringement and protects the patent from minor improvements; it protects the patent from imitators that try to invent around it. The greater is the number of equivalents assigned to a patent, the greater is the breadth of protection granted to it. Pioneer/drastic innovations are usually granted a greater range of equivalents than non-drastic ones (Merges and Nelson 1990, Miller and Davis 1990). In this way, the innovator of a pioneering invention is adequately rewarded with a broad patent.

The doctrine of equivalents is limited by the use of the doctrine of file wrapper estoppel (also known as the doctrine of prosecution history estoppel) (Miller and Davis 1990). The doctrine of file wrapper estoppel binds the inventor to the decisions that are made during the examination of the patent application. According to this doctrine, the inventor cannot resurrect, by the use of the doctrine of equivalents, subject matter surrendered during the patent examination. Thus, if the inventor agreed to amend his/her

---

[17] As in Merges and Nelson (1990) the term device refers to a product, a process or a compound.

claims (i.e., narrow them) in order to secure a grant (i.e., to avoid an objection by the examiner or to further support nonobviousness and novelty) (s)he is not allowed later to broaden the claims using the doctrine of equivalents. By broadening his/her claims the innovator is in effect canceling the amendments made before the Patent Office (Cornish 1989).

The doctrine of equivalents, by broadening the scope of protection beyond the boundaries of the claims, is a means of protection for the patentee against infringement. The doctrine of reverse equivalents and blocking patents are means of protection for alleged infringers (Merges and Nelson 1990).

The doctrine of reverse equivalents broadens the possibilities of infringement by not finding infringement even when the allegedly infringing device falls within the claims of the patent. If the accused infringer proves that his/her device is *substantially* different to the patented one in function, in the result that it accomplishes and in the way that it accomplishes the result, then infringement may not be found. The doctrine of reverse equivalents protects inventors of substantial technological improvements from infringement liability. The purpose of the doctrine of reverse equivalents, as it will become evident below, is mainly to prevent the patentee from using his/her patent to create a hold up problem.

In many cases the alleged infringer is a holder of a patent that represents an improvement of a broad patent. In litigation, the court may find the alleged infringer's patent to be valid (novel, useful and nonobvious and enabling) but also to infringe on the broad patent. In this case, the court in effect makes the alleged infringer's patent subservient to the broad, dominant patent (Merges and Nelson 1990).[18] This result leads to blocking patents. The holder of the subservient patent cannot use the improvement without a license from the holder of the dominant patent. At the same time, the holder of the dominant patent cannot use the improvement without a license from the holder of the subservient patent. Blocking patents usually result in cross licensing agreements between the patent holders.

Even though an inventor would rather not have his/her patent characterized subservient to a dominant patent, (s)he is better off with the patent being subservient than with no patent at all. A subservient patent gives bargaining power to its holder (Merges and Nelson 1990). The greater the value of the improvement, the greater the bargaining power of the subservient patent holder since the greater is the incentive for the dominant patent holder to cross license and use the improvement. On the other hand, the greater is the value of the improvement introduced by the allegedly infringing patent, the greater is the hold up power of the holder of the dominant patent. Since the subservient patent is not operative without a license from the holder of the dominant patent, the latter may use his/her advantage to extract as much innovation rent as possible from the subservient patent. In cases where the value of the technological improvement is substantial, the application of the doctrine of reverse equivalents can remove the hold up problem.

Merges and Nelson (1990) state that the choice of the doctrine applied by the courts should be based on the importance and the stand-alone value of the inventions

---

[18] The Patent Office grants improvement patents that are subservient patents to a dominant patent to which they represent an improvement. The claims in an improvement patent must be drafted in a 'Jepson claims' format. The Jepson claims have the same effect as a court ruling that a patent is subservient to a dominant patent (Merges and Nelson 1990). Inventors usually apply for 'independent' rather than for improvement patents.

that are litigated.[19] If the first patent protects a drastic innovation while the challenger's patent or invention is a minor improvement, then the doctrine of equivalents should be applied. If, however, the challenger's invention has a large stand-alone value as well, then the doctrine of blocking patents or reverse equivalents should be used. The doctrine of reverse equivalence should be applied in cases where the incumbent's patent has small stand alone value compared to the challenger's patent.

2.3.2 Patent Challenge and Invalidation

The validity of a patent may be challenged in two different ways – either directly by launching an opposition in the responsible Patent Office and/or in the courts or indirectly in the course of legal action against infringement. Attacking patent validity through the Patent Office rather than the courts has the advantage of being a quicker and less expensive way to oppose the granting of a patent. It should be viewed as an alternative, rather than an intermediate, way to opposing through the courts (Cornish 1989). The validity of a patent may also be challenged in the course of enforcing the patent holder's rights either in the form of defense assumed by the accused infringer or by means of a counterclaim for revocation. Irrespective of the way that patent validity is attacked the outcome of the challenge is determined by the patent claims.

Any person has the right to oppose the granting of a patent to the responsible

[19] Merges and Nelson (1990) provide excellent examples of litigated patents where the above doctrines where applied.

Patent Office within nine months of the grant in the case of the EPO and CPO[20] or anytime during the statutory life of the patent in the case of USPTO. The challenger does not have to establish a special interest to oppose (i.e., to be sued or to be threatened with an infringement suit) (Cornish 1989). Opposition is a means used to obtain the limitation or invalidation of a wrongly granted patent (EPO 2000a). The opposition is examined by an examination committee formed by technically and legally qualified examiners. If the opposition is filed before the granting of the patent, it will not be treated as an opposition but rather as an observation made by a third party, which will be taken into consideration by the patent examiner (see sub-section 2.2.3.2).

The opposition must clearly present the grounds for objection. These must fall in the following three categories: unpatentable subject matter, inadequate disclosure and unallowable amendments (Cornish 1989). The challenger may claim unpatentable subject matter on the basis that the invention is not novel, is obvious, has no industrial applicability (i.e., not useful) or is not regarded as an invention under the patent law. The opposition claim for inadequate disclosure should demonstrate that the way the invention is described in the specification does not enable a person skilled in the relevant art to reproduce it.[21] Finally, the challenger may claim that some of the rules were not followed during patent examination including the way the claims were drafted and the

---

[20] After the expiration of the nine-month period after the grant, any person who can prove that (s)he has been threatened with an infringement suit by the patent holder or that (s)he has instituted proceedings for a court ruling that (s)he is not infringing the patent may intervene in the opposition proceedings. Intervention may take place as long as the opposition is pending. The intervention will then be treated as an opposition (Cornish 1989).

[21] The challenger may provide as a proof of nonenablement experiments that he conducted that suggest that, based on the patent specification, he couldn't achieve the desired technical results (EPO 2000a). In this case the burden of proof falls to the patent holder who has to prove that the specification is enabling.

amendments were made.[22]

If the request for opposition is granted by the Patent Office (i.e., the opposition filed is found admissible) the patent will be reexamined. Reexamination usually concerns only the grounds of opposition brought forward by the patent challenger. The patent holder will be asked to file his/her observations with respect to the objections raised by the challenger (EPO 2000a). The patent holder may amend his/her application in reply to the grounds of opposition, or may try to provide convincing arguments as to why the patent should be held valid as it stands. The challenger will be asked to comment on the patent holder's response to the opposition. The examining committee, in addition to the written arguments/observations of the parties involved in the challenge, may request oral proceedings to clarify outstanding questions.

The reexamination of the challenged patent may result in the invalidation of the patent, in the maintenance of the patent in an amended form or in the rejection of the opposition. If the patent holder does not approve the amendments proposed by the Patent Office and/or does not submit an amended text, the patent will be revoked. If the patent holder provides amendments that do not satisfy the examining committee, the patent will also be revoked. If the amendments made with respect to the opposition grounds are found to be valid, then the patent will be maintained in the amended form. If the reexamination of the patent does not find any grounds for objections then the opposition will be rejected. The challenger cannot further appeal the Patent Office's decision to the courts. The only party which is still entitled to an appeal is the defendant.

---

[22] Amendments should not result in the extension of the subject matter beyond the content of the application as originally filed or in the extension of the protection granted by the patent (EPO 2000a).

Statistics from the EPO show that during the period 1978-1998, on average, 8.5 percent of the patents granted by the EPO were subject to opposition (Harhoff and Reitzig 2000). Harhoff and Reitzig (2000) state that the opposition procedures take on average 2.2 years if the patent is revoked and about 4 years if the scope of the patent is amended. Statistics from the USPTO show that less than one quarter of the opposed reexamined patents survive without any change (Barton 2000).

As mentioned above, direct validity challenges may also be launched in the courts. Any person who regards the patent to be invalid and wants to use the patented invention can directly ask the court for a declaratory judgment that the patent is invalid (Miller and Davis 1990). When the validity of the patent is attacked in the courts, the opponent must demonstrate that a controversy about the patent exists for the court to allow the suit. During trial the opponent has to present his/her case and try to prove that the patent is invalid to avoid liability, just as an alleged infringer would have to do (Miller and Davis 1990). The grounds of objection are the same as those on which an attack is initiated in the Patent Office.[23] The court may uphold the validity of the patent, it may revoke it or it may narrow its scope (by eliminating the opposed claims).

The courts allow questions of infringement and validity to be tried together since the defense of the accused infringer usually consists of claims that the patent is invalid

[23] Merges and Nelson (1990) give an example of an attack on the validity of the patent to the courts on the grounds of inadequate disclosure. They refer to a Supreme Court challenge brought by Thomas Edison in 1895 to a patent held by Sawyer and Mann for materials used in light bulb filaments. The patentees held a very broad patent, which claimed all carbonized fibrous, or textile material as incandescing conductors. Edison proved that the patentees had only proven that carbonized paper works as a conductor in light bulbs not all the material they claimed. He showed that the patent specification was nonenabling, since it did not specify which of the materials could be used as conductors and in which way. He showed that it took a lot of experimentation to demonstrate that a variety of a bamboo plant worked well as a conductor, a discovery that was not enabled by the patent specification. The court found the patent to be nonenabling and invalidated it. The court stated that the patent would have been upheld if the patentees had claimed only what they had invented.

and that it should be revoked. The accused infringer may use the doctrine of undue experimentation to prove nonenablement and thus noninfringement. The doctrine of undue experimentation is used to prove that the challenger had to experiment to produce the allegedly infringing device; thus the patent specification did not make it any easier to produce his/her device. If the alleged infringer can prove that the desired technical results of the patent could only be reached through experimentation then the specification is proven to be nonenabling, infringement will not be found and the patent will be revoked (for an example see footnote 24).

It is not unusual during an infringement lawsuit for a patent to be found invalid and to be revoked. Miller and Davis (1990) claim that alleged infringers have been successful in attacking patents, as US courts have found litigated patents invalid more often than valid. Choi (1998) finds that only thirty per cent of the patents contested during an infringement lawsuit during the period 1966 to 1971 where found to be valid.

It may take more than one court ruling to determine the technological territory protected by the patent, if any. The courts have discretion when they judge questions on patent infringement and validity. Thus, it is not unusual for a court decision to cancel another court decision. Merges and Nelson (1990) provide examples of court rulings that overruled previous court rulings (for an example see footnote 5).

The discretion allowed to both the Patent Office and the courts is due to the fact that the claims cannot be pre-specified and standardized. The breadth of patent protection, which is one of the most important parts of the patent, is often determined after the patent is granted by the way the patent claims are interpreted during a patent validity challenge and/or an infringement trial. Thus, the patentee needs to be careful when (s)he drafts his/her claims. The claims must provide protection against those looking for ways of circumventing the patent and at the same time, they must not break any of the validity rules. Thus, the claims should cover all possible embodiments of the invention while avoiding to claim embodiments that are anticipated or obvious. Cornish (1989) cautions patent applicants about the hazards of the patent system. He describes it as a game in which only the very skilled can be successful and emphasizes the need of getting everything right from the beginning.

2.4 Patent Breadth and Innovation Rents

The previous sections of this chapter examined the innovator's role in the patent granting process and in the determination of the breadth of protection covered by his/her patent. The various ways a patent may be contested after it is granted have also been considered. The innovator's patent claims are crucial in determining the outcome of the patent granting process (i.e., whether the patent will be granted and the breadth of patent protection) and the ability of the innovator to safeguard and/or defend his/her patent after it is granted. The present section examines the importance of patent breadth in determining the level of innovation rents that the patentee can appropriate with his/her patent.

Patent breadth determines the degree of competition allowed in the market and the effective patent life, which in turn determine the reward to the innovator. The standard argument in the economic literature on patent breadth is that the patent applicant will try to maximize the rents appropriated with his/her innovation by claiming the broadest scope of patent protection possible (Lanjouw and Schankerman 2001, Merges and Nelson 1990, Gilbert and Shapiro 1990).

The above argument is mainly based on two assumptions. The first assumption holds that a positive linear relationship can be assigned between patent breadth and the flow of profits to the innovator (Gilbert and Shapiro 1990, Gallini 1992). The second assumption is that the Patent Office will prune back or reject broad claims and will determine a socially optimal breadth of patent protection. If the first assumption holds, then a profit-maximizing innovator should apply for the broadest protection possible. If the second assumption is true, then the optimal patent breadth will always be determined by the Patent Office. The innovator in this case should apply for the broadest protection, since the breadth of protection granted cannot be greater than the breadth of protection claimed. In other words, if the innovator applies for narrow protection the Patent Office cannot extend his claims and grant him/her broad protection even if that would have been optimal.

Regarding the first assumption, it is generally true that a patent with a very narrow protection (zero breadth) can be of no use to the patentee as competitors with almost identical inventions can enter his/her market.[24] This is the type of protection offered in the early days of the patent system. Klemperer (1990) refers to the example of the cotton gin first patented by Eli Whitney in 1794 which offered the inventor practically no protection since his competitors could get patents on almost identical inventions.

It is also generally true that the broader the protection granted by the patent, the harder it is, *ceteris paribus*, for a new technology to render the existing one obsolete without infringing the patent. Thus, broader patents may be associated with a longer

---

[24] In the present context zero breadth means protection only against duplication of the patented innovation.

effective patent life. In addition, the broader is the patent protection, the less competitive is the market expected to be, since the harder it is for a competitor to enter without infringing the patent. Given that the longer is the effective patent life and the smaller is the number of competitors that can successfully enter the patentee's market the greater are the profits for the patentee, it is a reasonable conclusion that there is a positive relationship between patent breadth and patent profits.

The above argument holds true as long as the patent, once granted, is not challenged legally. This has not been the case, however. As it has been demonstrated in Section 2.3, a patent may be contested either when the patented subject matter is infringed or when its validity is directly challenged. The role of the courts in refining patent scope has been crucial as they have often invalidated or made a ruling that effectively narrowed the scope of a broad patent (Merges and Nelson 1990, Miller and Davis 1990). In fact, the broader is the patent protection, the higher is the probability that the patent will be challenged legally, that it will overlap another patent and/or that the courts will invalidate it or narrow its scope (Lerner 1994, Lanjouw and Schankerman 2001). Given the above, a broader patent breadth may shorten the effective patent life, which in turn means that a positive linear relationship between patent breadth and patent profits cannot be assumed.

The assumption that the Patent Office will always eliminate broad claims and will determine a socially optimal scope of protection is also not realistic. Barton (2000) points to resource limitations at the USPTO generated by the increase in patent applications over the last decade, which induce the patent examiner to often side with the applicant and grant broad claims. The inefficiencies inherent in the operation of the Patent Office can often result in the granting of broad patents that cannot survive a validity attack and to patents that overlap leading to disputes that have to be resolved through costly litigation or settlement.

Despite the above-mentioned problems related to patent breadth, broad patents are observed in practice, especially in biotechnology, which implies that the innovator has applied for broad patent protection. This is mainly due to the fact that biotechnology is an area that generates many drastic innovations, which as was previously mentioned are usually granted broader protection. Lentz (1988, p.318) also refers to the myopic behavior of the innovator when he concludes that broad biotechnology patents are observed due to the 'inherent imprecision of the biological studies, overreaching and/or naivete.' Another explanation could be that the innovator relies on the Patent Office to structure his claims. Given the existing inefficiencies in the operation of the Patent Office, which leads to the granting of broad patents and the fact that broad patents are often overturned, patentees need to carefully consider the breadth of the patent protection that they are claiming.

2.5 Concluding Remarks

This chapter examined the evolution of the patent system, the patent granting process, the different ways a patent may be contested after it is granted and the economic importance of patent breadth. The innovator's role in determining whether the patent will be granted by the Patent Office and whether it will be viable after the grant was also examined in this chapter.

The innovator through his/her claims in the patent application defines the technological territory over which protection is sought. This territory represents the breadth of patent protection claimed. The Patent Office examines the claims made by the innovator and decides on the breadth of patent protection granted to the patent, if any. The greater the number of independent claims made (and upheld by the Patent Office) and the more general are the claims, the greater is the breadth of patent protection claimed (and granted).

The innovator would like to draft his/her claims so as to safeguard the broadest protection possible for his/her innovation. At the same time, however, the claims must satisfy the patentability requirements of novelty, nonobviousness, utility and enablement. Thus, on the one hand, broad patent protection is desirable to the innovator as the broader is the patent, the harder it is for competitors to enter the patentee's market and the greater are the profits that the innovator can expect to appropriate with the patent. On the other hand, the broader is the patent protection claimed, the harder it is for the innovator to secure a patent grant, to protect the patent from infringement and to defend the validity of the patent. That is, the broader is the patent protection claimed, the harder it is to differentiate the innovation from the prior art (i.e., to show novelty), to demonstrate that the innovation is not anticipated or obvious and to provide an enabling disclosure. In addition, even when broad claims are granted by the Patent Office (which is usually the case for drastic innovations), the broader is the patent protection, the greater is the probability that the patent will be infringed, that it will overlap another patent and that the courts or the Patent Office will invalidate it or narrow its scope.

This chapter has demonstrated that the innovator plays a crucial role in determining both the patentability of his/her innovation and the viability of the patent after the grant. It also became evident in this chapter that a strong and viable patent is not necessarily a broad patent. The chapter that follows reviews the existing patent literature and explores how the patent breadth issue has been treated in the various economic studies.

CHAPTER III

THE ECONOMICS OF PATENT SCOPE

3.1 Introduction

The economic literature on the patent system is substantial. Early economic studies focused on the welfare implications of patents and examined whether the patent system should be retained or abolished. A large number of studies that followed dealt with issues concerning the efficiency of the patent system as a second-best solution in promoting the production of knowledge and rewarding the innovator. The durability of the patent system as a feasible alternative to generate knowledge led economists to try to fine tune its different aspects and reduce its inefficiencies. This work led to studies that focused on the design of an optimal patent policy using a variety of policy tools. Initial studies examined patent length and compulsory licensing, while recent studies have focused on the economic importance of patent breadth and its use as a policy tool in designing a socially optimal patent protection. The increase in litigation during the last decades has initiated studies that empirically examined the factors that expose patents to litigation risk.

This chapter reviews the economic literature on patents and focuses on studies that examined the economic importance of patent breadth, its use as a policy tool and its effect on the pace of innovation. The rest of the chapter is organized as follows. Section 3.2 reviews the studies that focused on the welfare implications of the patent system and its effectiveness in stimulating technical progress and in rewarding the innovator. Section 3.3 discusses the findings of studies that examined the determination of an optimal patent policy. The use of patent breadth as a policy tool in designing an optimal patent policy is discussed in section 3.3. Section 3.4 reviews the studies that examined patent litigation and patent opposition. Finally, section 3.5 discusses how the innovator's patent breadth decision has been treated in existing studies and concludes the chapter.

3.2 The Impact of Patents on Welfare and Innovation Appropriability

Studies of the welfare effects of patents typically use one of two benchmarks to compare the pre-patent to the post-patent outcome. One benchmark is that, in the absence of the patent system, the innovation would not have taken place. The other benchmark is that, in the absence of the patent system, the invention would have been provided by the public sector at marginal cost. These two benchmarks generate very different welfare outcomes. As Machlup (1958) notes, studies using the former benchmark concluded that there is no loss in consumer surplus because of the patent system while studies that used the latter benchmark typically concluded that the patent system generates deadweight welfare losses. Both the above benchmarks have been used in the recent literature. For instance, Nordhaus (1969) and Tandon (1982) have assumed that in the absence of the patent system, the innovation would not have taken place, while Klemperer (1990) assumed that public provision of the innovation would have occurred.

A review of the early literature on the efficiency of patents in promoting technical change shows that economists were rather agnostic on whether patents in fact promoted technical progress. Machlup (1958) indicates that he could neither prove nor disprove that patents were promoting technical progress based on the then existing empirical and theoretical findings. Jewkes et al. (1960) noted that the patent system lacked logic, and was crude and inconsistent leading to extensive and costly litigation. Despite these problems, however, they suggested that in the absence of a better alternative the patent system should be retained as a system of stimulating the production of knowledge. Nordhaus (1969) compared patents to subsidies with respect to their effectiveness in promoting technical change and was unable to conclude that one was superior to the other. He nevertheless found the patent system to be 'a remarkable device' (Nordhaus 1969, p.90).

Empirical studies that followed shed some light on the role of patents in stimulating the production of knowledge and on their efficiency in rewarding the innovator. Mansfield et al. (1981) studied the relationship between patents and the rate of innovation. They found that the effect of the patent system on knowledge generation differed depending on the industry studied. Their results showed that twenty-five (various industries) to fifty percent (drug industry) of the innovations would not have taken place in the absence of the patent system. In this study they also examined the influence of patents on imitation costs. They found that patents increase imitation costs but do not make entry impossible or unlikely. They found that sixty per cent of the patent innovations that they studied were imitated within four years of their introduction.

Levin et al. (1987) empirically studied the appropriability of the returns from industrial R&D in more than one hundred manufacturing industries. They found that patent protection raised imitation costs by almost forty per cent for new drugs, about thirty per cent for new chemical products and twenty five per cent for typical chemical products. They also found that not all innovations were patented.

Schankerman (1998) empirically studied the effect of patents on the R&D incentive by measuring the returns to R&D using patent renewal data in France for different technology fields. He concluded that the patent system created valuable property rights that were equivalent to an R&D cash subsidy rate of fifteen to twenty five per cent (average across technology fields). He also found that R&D incentives were shaped not only by the patent law but also by institutional constraints (i.e. price regulations).

Trajtenberg (1990) used both patent counts and patent citations as instruments to empirically determine the economic value of innovations. He concluded that patent counts were highly correlated with concurrent R&D but they were not indicative of the value of the innovation. He found a positive correlation between patent citations and the social value of the innovations cited.[1]

Lerner (1994) provided an empirical analysis of the importance of patent scope for the value of a firm. He studied biotechnology firms and used as a proxy for patent scope the number of classes in the International Patent Classification system that the patent examiner assigned to each patent. He found that broad patent protection increased the value of the firm significantly and that the greater was the number of substitutes in a product class, the greater was the impact of patent breadth on the value of the patent.

---

[1] In this study, the social value of innovations is measured by $\Delta W_t = W(S_t) - W(S_{t-1})$ where $W(\cdot)$ is an estimated social surplus function and S_{t-1} and S_t a set of products offered in two successive periods; the period before (t-1) and after (t) the introduction of the innovation.

Gilbert and Newbery (1982) theoretically examined the efficiency of preemptive patenting (i.e., patenting innovations that are not used or licensed) as a tool used by an existing monopoly to maintain its monopoly power. They found that under certain conditions (e.g., under a perfect market for R&D inputs) preemptive patenting could enable a firm to maintain a monopoly position despite the potential of entry.

Waterson (1990) theoretically examined the ways that a product patent could change the nature of market entry behavior. He found that patents did not make entry impossible, a result that supports the empirical findings of Levin et al. (1987) and Mansfield et al. (1981). He also found that patents could lead to a better location of products in the economic space and that under certain conditions social welfare could be enhanced if patents could deter entry. In this study, Waterson also dealt with the innovator's decision to patent the product or to keep it a secret when the innovator faces possible court action. He concluded that not all innovations would be patented (supporting the findings of Levin et al. (1987)) even though the innovations that were not patented would often be copied.

Horstmann et al. (1985) examined the patenting behavior of an innovating firm (the decision to patent the innovation or to keep it a secret) when the patent reveals information that is of importance to competitors. Their findings confirm Waterson's findings that not all innovations will be patented. However, their model suggested that the innovations that were not patented would be imitated rather than copied. They also found that the innovator will always find it optimal to patent the innovation when competitors find imitation unprofitable.

Other studies examined the use of price discrimination and licensing as means of enhancing the appropriability of a patented innovation. Hausman and Mason (1988) analyzed the relationship between price discrimination and patent policy. They suggested that price discrimination would allow innovators to appropriate more of the benefits from their innovation, thus leading to innovations that otherwise would not have been produced. They concluded that third degree price discrimination by patent holders could raise social welfare.

Kamien and Tauman (1986) studied the effect on the innovator's returns of licensing a patented cost-reducing innovation in an oligopolistic market structure. They compared licensing by means of a fixed fee to licensing by means of a per unit royalty. They found that a fixed fee results in greater returns for the patentee and is preferred by consumers because it results in lower prices compared to licensing by means of a per unit royalty. They also found, however, that a fixed fee might affect the market structure while a per unit royalty would not affect the final number of firms in the market.

Gallini and Wright (1990) studied licensing under asymmetric information and determined the form of a licensing contract (exclusive versus non-exclusive and fixed fees versus output-based royalty) that would maximize the returns to the innovator. In their model, the innovator has private information on the value of the innovation and the transfer of information facilitates imitation. They found that low value innovations can be fully exploited with an exclusive contract and a fixed fee. When the value of the innovation is high, however, the innovator will offer an output-based royalty and the degree of exclusivity of the contract will depend on the level of imitation costs.

Rockett (1990) demonstrated how licensing could be used by a patent holder as a means of choosing the nature of competition (s)he would face when his/her patent expired. She found that licensing the patent to 'weak' competitors (in terms of size or marginal costs) is an effective strategy that allows the patentee to prolong his/her dominant position in the market after the patent expires.

3.3 Optimal Patent Policy and Patent Breadth

The economic studies considered so far focused on the welfare effects of the patent system, or its efficiency in generating technical knowledge and in rewarding the innovator. The studies examined in this section focus on correcting the inefficiencies associated with the patent system through the determination of a socially optimal patent policy. A variety of policy tools has been used in the optimal patent policy literature, e.g., patent length, compulsory licensing and patent breadth. Patent length was the first policy tool examined.

Nordhaus' (1969) pioneering work examined the socially optimal patent length for a cost reducing process innovation. He concluded that the socially optimal patent length should be finite to reduce monopoly inefficiencies despite the fact that a shorter patent life might hinder research. Scherer (1972) offered a geometric reinterpretation of Nordhaus's optimal patent length work supporting his findings.

Tandon (1982) used both patent length and compulsory licensing as policy instruments in designing a socially optimal patent protection. He suggested that the use of compulsory licensing would combat the deadweight welfare losses of the patent system.[2] He found that with compulsory licensing the optimal patent life could be infinite for both process and product innovations. Scherer (1980) suggested that the

---

[2] For both Tandon (1982) and Nordhaus (1969) the deadweight welfare losses are associated with longer versus shorter patents since as mentioned earlier both studies assumed that in the absence of patent protection the innovation would not have take place.

application of compulsory licensing to cases where patent-based monopoly power had been abused would have no adverse impact on the rate of technological progress, supporting the findings of Tandon (1982) that compulsory licensing could be welfare enhancing. Merges and Nelson (1990) concluded, however, that despite the theoretical proofs of the merits of compulsory licensing, this policy instrument has been rejected by the intellectual property community and thus its use as a tool to reduce the monopoly inefficiencies of the patent system seems unlikely.

In the early 1990s, economists started using another dimension of the patent – patent breadth – in designing optimal patent policy. Merges and Nelson (1990) provided an excellent study of patent breadth and demonstrated, using patent cases, the importance of patent breadth for the level of competition in the market, the returns to the patentee and the viability of the patent after it is granted.

While patent length is easily defined and agreed upon, the meaning of patent breadth is less clear-cut. Patent breadth has been defined in terms of the area in the horizontally differentiated (Klemperer 1990) and/or in the vertically differentiated (O'Donoghue et al. 1998, O'Donoghue 1998) product space that the patent protects. These are areas in product space in which a competitor cannot offer a product without infringing the patent. The terms lagging breadth and leading breadth have also been used to denote the breadth of patent protection (O'Donoghue et al. 1998, O'Donoghue 1998). Lagging breadth refers to a set of inferior products, for which no further innovation is required for their production; however, if these products are produced they would infringe a patent. Leading breadth refers to a set of products that require further innovation to be produced; if produced they too would infringe a patent. Both lagging and leading breadth can refer to either the horizontally or the vertically differentiated product space. Patent breadth has also been defined in economic studies in terms of how strong the novelty requirement should be for a subsequent innovation not to infringe the patent (Scotchmer and Green 1990, Van Dijk 1996). The difference in the economic definitions of patent breadth is in part responsible for the conflicting findings of the various studies regarding the optimal patent shape (i.e., patent breadth and patent length). These findings will now be reviewed.

Gilbert and Shapiro (1990) and Klemperer (1990) were among the first studies to use both patent length and patent breadth in the design of an optimal patent policy. Gilbert and Shapiro (1990) studied a homogenous good market and determined the optimal mix of patent breadth and patent length (using them as substitutes) that would guarantee a given reward to the innovator at the minimum social costs. They defined patent breadth in terms of the flow of profits that it confers to the innovator assuming a positive linear relationship between the two. In their model, greater breadth leads to a greater flow of profits but also to greater deadweight welfare losses. They concluded that optimal patent policy calls for narrow but infinite lived patents.

Klemperer (1990) determined the optimal patent shape (patent breadth/width and patent length) for a horizontally differentiated product innovation that would confer a given reward to the innovator at the least social cost. He defined patent breadth in terms of the area in the horizontally differentiated product space that the patent protects. He assumed that patented varieties of a product are preferred to the non-patented ones by all consumers, but consumers differ in their demands and costs of substituting to less preferred varieties. In his model the design of optimal patent policy depends on consumer characteristics. When all consumers face the same cost of switching to less preferred varieties infinite lived and narrow patents were found to be optimal. If all consumers have the same cost of substituting outside the product class, however, broad and short patents were found to be optimal.

Gallini (1992) determined optimal patent policy under costly imitation. She suggested that the longer is the patent life, the greater the incentive to imitate the patent. Defining patent breadth as the flow rate of profits for the innovator and using as instruments of patent scope either the cost of imitation or the flow of profits to the imitators she finds that optimal patents are short and broad. Her results contrast the previous studies of Tandon (1982), Gilbert and Shapiro (1990) and in part Klemperer (1990) which suggest that infinite lived and narrow patents are optimal.

The main reasons behind the difference in the results on the optimal shape of a patent (i.e., broad and short versus narrow and long) are the way patent breadth is defined and imitation costs are treated in these studies. Tandon (1982), Gilbert and Shapiro (1990) and Klemperer (1990) assume that imitation is costless and therefore always a threat to the innovator, while Gallini (1992) assumes that imitation is costly but not prohibitively so.

Eswaran and Gallini (1996) examined the role of patent policy in directing technological change by determining the socially efficient mix of process and product innovations. They found that the private solution (i.e., absence of government intervention) leads to too many product innovations (that relaxes competition through product differentiation) relative to process innovations (that reinforces competition through the reduction in production costs). To deal with this problem they proposed broad process and narrow product breadths when the R&D costs of the pioneer are low (i.e., enhance competition by forcing the entrant to develop a more efficient process and thus become a fiercer competitor), and broad product and narrow process breadths when the R&D costs of the pioneer are high (i.e., dampen competition by forcing the entrant to develop a very different product).

The above studies examined patent breadth in a context of a single innovation ignoring the effect of patent breadth on subsequent innovations. In addition, these studies do not determine the magnitude of the return to the innovator; thus they do not deal with the question of how to preserve the incentive to innovate by efficiently dividing profits between sequential innovators. These questions are addressed by a group of studies that is discussed below.

Scotchmer and Green (1990) studied how the novelty requirement affects the pace of innovation. They defined patent breadth in terms of how strong the novelty requirement must be for a subsequent innovation not to infringe the initial one. They used a dynamic model of a patent race to describe how the strength of the novelty requirement and the rule by which a patent is awarded (first to invent versus first to file) determine the decisions to patent and to enter or exit a patent race. They found that a weak novelty requirement allows for an intermediate patent increasing social welfare from early information disclosure, while a strong novelty requirement allows only for a final patent. They concluded that the stronger the novelty requirement (i.e., the greater the patent breadth), the greater the profits for the innovator, the greater the loss in welfare from limiting information disclosure and the higher the incentive for firms to drop out of a patent race when they fall behind technologically.

O'Donoghue et al. (1998) studied the effect of patent life, leading and lagging patent breadth on the pace of technological progress. They found that patent breadth in general and leading breadth in particular can increase the rate of innovation. In their model, a given rate of innovation can be achieved either through a patent of infinite life and narrow leading breadth or through a patent of infinite leading breadth and short life. They found that the former patent shape is more efficient in minimizing R&D costs and the latter is more efficient in minimizing the costs of delayed diffusion.

Van Dijk (1996) studied the role of the novelty requirement once the patent is granted on the competition for product improvements (i.e., vertical differentiations of a basic product). He defined patent breadth (which he terms patent height) in terms of the stringency of the novelty requirement that the patent examiners use. He found that low patent protection does not affect the market equilibrium without a patent (i.e., low patent protection does not affect the level of competition in the market). He also found that intermediate patent protection might even hurt the patent holder of the basic innovation as the non-patent holder may make a credible commitment to choose the most profitable improvement (greater height/breadth). Finally, he concluded that broad patent protection will benefit the patentee.

Scotchmer (1991) studied the division of profits between innovators when one innovator's technology builds on another's. She did not determine an optimal patent breadth; she rather described how patent breadth affects the incentive to innovate first and second generation products. She found that cooperation among innovators (research joint ventures) and patent breadth (that determines the bargaining power of the parties) can protect the incentive to innovate by guaranteeing an efficient sharing of profits. She suggested that when research joint ventures are not possible, courts should allow collusive licensing between competing patentees to prevent erosion of their profits by competing in the market.

Chang (1995) also studied the effect of patent breadth on cumulative innovation determining the optimal court rulings that would lead to an efficient allocation of profits between first and second-generation innovators. He found that, as is suggested by Merges and Nelson (1990), broad protection should be granted to innovations with large stand-alone value relative to subsequent improvements. As well, he found that innovations with very small stand-alone value relative to improvements, when the former are building stones for the latter, should also be offered broad protection. This kind of protection guarantees that there are enough incentives for the generation of the first innovation. He suggested that collusive agreements between patentees could lead to inefficient entry by imitators who invent around the original patent.

Green and Scotchmer (1995) showed that an efficient division of profits between sequential innovators depends on patent breadth and the opportunity for cost sharing through *ex ante* and/or *ex post* licensing. They used patent breadth and patent length as complementary tools rather than substitutes. In their model, patent breadth determines the relative bargaining power, while patent length determines the magnitude of joint profits. They found a positive relationship between the breadth of the initial innovation and the speed of disclosure. They also found that under certain licensing conditions a narrow patent can leave the first innovator better off than can a broad one.

O'Donoghue (1998) studied sequential innovations where firms repeatedly supersede each other. He identified four instruments of patent policy – the life of the patent, lagging breadth, leading breadth, and the patentability requirement. He focused on the latter and proposed the requirement of a minimum innovation size for patents. He found that the requirement of larger innovations would stimulate R&D since doing so is equivalent to requiring future innovators to achieve larger innovations thus extending the effective patent life.

Matutes et al. (1996) determined the optimal patent policy when the initial innovation is fundamental and does not have a commercial value. They determined breadth as the number of potential applications that can be granted to an innovator of a fundamental idea or process. They found that the breadth of the initial innovation and the speed of disclosure are positively related, a result that contradicts the findings of Scotchmer and Green (1990). This different result, according to the authors, is due to the different definition of the breadth of patent protection. In Matutes et al. (1996) model patent breadth is defined in terms of the courts' interpretation of the patent claims, while in Scotchmer and Green's (1990) model patent breadth is defined in terms of the courts' interpretation of the novelty requirement.

Scotchmer (1996) also studied the protection of a fundamental innovation and investigated whether second generation products and applications of the innovation should be protected through exclusive licenses or through patents when they infringe the fundamental innovation.[3] She found that if *ex ante* exclusive licenses are permitted then the applications or the improvements of the fundamental innovation that infringe it should not be patentable. If the improvements are not patentable, the patent holder of the fundamental innovation may appropriate more profits of the sequential innovations.

3.4 Patent Opposition and Litigation Risk

The escalation of patent oppositions and litigation over the last two decades initiated studies that examine the factors that expose patents to opposition and litigation risk.

[3] The status of improvements and applications (new uses) of an innovation regarding infringement and patentability is ambiguous under the patent law (Scotchmer 1996). Thus, an improvement or an application of an innovation may infringe the basic innovation but it may also be patentable (see discussion on subservient patents in sub-section 2.3.1, chapter II).

Harhoff and Reitzig (2000) empirically examined the determinants of opposition against EPO patent grants. They used a data set covering all biotechnology and pharmaceutical patents granted by the EPO during the period 1979-1996. They found that the perceived value of the patent (measured by the number of backward and forward citations and by the number of countries in which protection is sought) increased the probability of opposition. They also found that the existence of information asymmetries between the parties involved concerning the value of the invention and the outcome of the opposition also increased the probability of opposition. They captured the above mentioned information asymmetries by comparing patents in new biotechnology areas (e.g., microorganisms, enzymes, recombinant DNA processes) to a control group of patents with low uncertainty (patenting of machinery).

Lanjouw and Schankerman (2001) empirically studied the factors that expose patents to litigation risk. They used a data set covering all patent validity and patent infringement suits involving patents granted by the USPTO during the period 1975-1991. They found the probability of litigation to vary significantly across patents depending on the technological area. They observed very low litigation rates for patents protecting chemical innovations and very high litigation rates for patents protecting pharmaceutical innovations (one case filed for every fifty drug patents). Among the main factors affecting the probability of litigation were, the type of ownership (corporate/individual), the nationality (US/foreign) of the owners, the backward and forward citations made and received by a patent, respectively, the forward self citations and the number of claims in the patent. They found that the value of the patent (measured by the number of forward and backward citations), the innovator's engagement in subsequent improvement innovations (measured by the number of forward self citations) and the number of patent claims increase the probability of both infringement and invalidity suits.

3.5 The Innovator's Patenting Decisions

The studies discussed above examined important aspects of patent protection as a means of appropriating innovation rents. The innovator's role in the patenting process, his/her choice of the breadth of patent protection claimed, once the decision to seek patent protection has been made, has not been explicitly examined in the above studies, however. Instead, it has been implicitly or explicitly assumed that the innovator has an incentive to claim as much as possible (Merges and Nelson 1990, Gilbert and Shapiro 1990, Lanjouw and Schankerman 2001).

As discussed in detail in chapter II, the innovator determines whether the patent will be granted and the viability of the patent after the grant through the patent claims that (s)he makes. Thus, the innovator's patenting behavior is crucial in determining the level of innovation rents that can be appropriated with the patent. The ability to appropriate innovation rents is, in turn, crucial for the incentive to invest in R&D and to generate the innovation.

Lerner (1995) is the only study – to the best of the author's knowledge – that examined some aspects of the innovator's patenting behavior. Lerner (1995) empirically examined the patenting behavior of 419 new biotechnology firms that differed with respect to their litigation costs. He used as proxies for litigation costs the number of previous patent suits in which the firm had been involved and the firm's financial resources. He found that the firm's decision concerning the subclasses in which patent protection should be sought depended on the firm's litigation costs and the litigation costs of the firm's rivals. He found that firms with high litigation costs were less likely to patent in the same subclasses as their rivals, especially if their rivals were firms with low litigation costs. He also found that firms with the highest litigation costs were twice as likely as others to patent in subclasses with no rivals.

Even though Lerner's (1995) empirical study shed some light on how innovators in the field of biotechnology choose certain characteristics of their patents, there is no formal model of the innovator's patent breadth decision in the patent literature. The only aspect of the innovator's patenting behavior that has been explicitly modeled is the decision to patent the innovation or to keep it a secret (Waterson 1990, Gallini 1992, Horstmann et al. 1985).

The present study extends the patent literature by explicitly modeling the patenting behavior of an innovator of a drastic product and a drastic process innovation. In specific, the theoretical models developed in chapters IV and V model the patenting behavior of an innovator who, having invented a patentable innovation and having decided to seek patent protection, decides on the breadth of patent protection claimed. The innovator in these models determines the optimal patent breadth claimed, that is, the breadth of patent protection that maximizes his/her ability to appropriate innovation rents when (s)he is faced with a positive probability of patent infringement and/or a direct validity attack. The models developed should be seen as a first step towards the study of the innovator's patent breadth decision.

CHAPTER IV

STRATEGIC PATENT BREADTH FOR DRASTIC PRODUCT INNOVATIONS

4.1 Introduction

This chapter models the patenting behavior of an innovator who, having invented a drastic product innovation, determines strategically the optimal breadth of patent protection. A strategically acting innovator takes into consideration the effect that patent claims have on potential competitors and realizes that the effort of safeguarding his technological territory does not conclude with the granting of the patent. The innovator takes into consideration the possibility of having to enforce and/or defend his patent rights when he determines the breadth of patent protection claimed. As a consequence, the model explicitly examines the assumption that the patentee always applies for the broadest scope of patent protection.

The chapter focuses on patents that involve drastic product innovations. Drastic product innovations are innovations that generate new demand or meet demand not previously met, thus enabling the innovator to exert monopoly power in the market. The focus is on drastic innovations for two reasons. First, when drastic innovations are introduced, the innovation rents that are at stake are substantial, thus increasing the probability of litigation. Second, according to the Patent Office's policy, drastic innovations are usually granted broader protection (EPO 2000a, USPTO 1999). This is mainly due to the fact that the patent examiner cannot find support from the prior act to object to broad claims. Thus, claims to drastic inventions are often allowed to cover areas beyond the area examined and disclosed by the inventor. Nelson and Merges (1990) observe that the narrowing of the claims of pioneering/drastic inventions is left to the courts. In addition, as discussed in chapter II, the greater is the breadth of patent protection, the greater is the probability that the courts will invalidate the patent during a patent infringement trial or a direct validity challenge. Thus, the patent applicant needs to be more careful when he drafts his claims when drastic innovations are concerned, since he cannot depend as much on the Patent Office for guidance in structuring his claims.

The drastic product innovations that are considered in this chapter take place within a vertically differentiated product space. Products are vertically differentiated when one product is preferred to the rest by all consumers if offered at the same price (Cremer and Thisse 1991). Consumers can thus rank the products in terms of a characteristic, usually quality. Although consumers have the same most preferred characteristic (quality), they have a different willingness to pay for the characteristic.[1]

The rest of the chapter is organized as follows. Section 4.2 describes the theoretical framework that is developed to examine the innovator's strategic patenting behavior. This section analyses the conditions that prevail in the market and defines

[1] In contrast, products are horizontally differentiated when consumers cannot rank them. That is, all products would have positive market shares if offered in a market at the same price (Beathe and Katsoulacos 1991). Under horizontal differentiation consumers have different most preferred varieties but they all have the same willingness to pay for their most preferred variety.

patent breadth. Section 4.3 presents the analytical solution to the strategic patent breadth game. This section first describes the entrant's location decision under no patent protection and under patent protection. Section 4.3 then examines the patentee's choice of the optimal patent breadth. Section 4.4 concludes the chapter.

4.2 Theoretical Development of the Strategic Patent Breadth Model

The patenting process is modeled as a sequential game of complete and perfect information. The agents involved in the game are an incumbent/patentee who applies for a patent and decides on the breadth of patent protection claimed, and a potential entrant who decides where to locate in the product space and who potentially competes with the incumbent in the market. It is assumed that the incumbent has invented a product that meets the patentability requirements and that the regulator (i.e., Patent Office) always grants the patent as claimed; thus the regulator is not explicitly modeled. The latter assumption is made to reflect the case under which the innovator has no help from the Patent Office in determining the breadth of patent protection. The assumption that the Patent Office grants the patent as claimed is a realistic assumption for drastic innovations.

The game consists of three stages. At the first stage of the game the incumbent, having invented a drastic product which will allow him to exert monopoly power in the market and having decided that he wants to patent it, determines the breadth (b) of protection that will be claimed. In the second stage an entrant observes the patentee's product and the breadth of protection granted to it and chooses whether to enter the market or not. If the entrant does not enter, the patentee operates as a monopolist in the third stage of the game. If the entrant decides to enter she does so by choosing the characteristics of her product – i.e., her location in the product space. Once the entrant enters, two cases may prevail in the third stage of the game. If the entrant does not locate within the patentee's claims or if she locates within the patentee's claims and infringement is not found then both the patentee and the entrant market their products and compete in prices. However, if the entrant locates within the patentee's claims and infringement is found then the entrant is not allowed to market her product and the patentee operates as a monopolist at the third stage of the game. It is assumed that both the patentee and the entrant are rational and foresighted. Thus, they both fully anticipate the reaction of their opponent to each of their actions.

In this model, the patentee determines the breadth of protection that will allow the maximum appropriation of innovation rents. The patentee acts strategically taking into consideration the entrant's responses to his choice of patent breadth. He is aware that the probability of the patent being attacked and/or invalidated increases with the breadth of protection and that a broad patent could impede his ability to enforce his rights if the patent is infringed and/or if its validity is directly challenged. In addition, the innovator does not rely on the Patent Office for guidance in structuring his claims. He is aware of the inefficiencies in the determination of patent breadth in the Patent Office and of the fact that his effort to safeguard his technological territory does not usually conclude with the granting of the patent.

The incumbent's decisions to invent (i.e., his choice of the innovation's specifications) and to patent are not examined – these decisions are treated as exogenous to this game. The only decision that the incumbent has to make involves the breadth of protection that will be claimed – the length of protection is predetermined and is the same for all patents. It is assumed that the incumbent's patent does not infringe on any previous product or process patent. It is also assumed that there is only one Patent Office in which the incumbent can apply for patent protection (or equivalently that Patent Offices do not differ with respect to the protection that they grant). In addition, it is assumed that the patentee and the entrant produce only one product each and that the entrant does not patent her product since further entry is not anticipated. It is also assumed that the production process is deterministic; once the entrant chooses a location she can produce the chosen product with certainty. Finally, it is assumed that there is no time lag between making and realizing a decision.

4.2.1 Market Conditions and Determination of Patent Breadth

The market in which the patentee and the entrant operate is characterized by the following conditions. The market can only sustain two products, that is, the market is a natural duopoly. Every product i (i=1, 2) has an inherent quality represented by the parameter q_i. The quality parameter q_i takes values in the interval $[\underline{q},\bar{q}]$ where $\underline{q}$ corresponds to the lowest quality that can be allowed in the market while $\bar{q}$ corresponds to the highest quality that is technologically feasible. All consumers agree that higher quality is preferred to lower quality. It is also assumed that every product can be completely described by its quality alone (i.e., only one variety of the product can be produced).

The market consists of differentiated consumers uniformly distributed in the interval [0, 1], each buying one unit of either product 1 or 2 but not both. The consumers differ with respect to some attribute λ uniformly distributed with unity density $f(\lambda)=1$

in the interval $\lambda \in [0, 1]$. The attribute λ determines a consumer's willingness to pay for quality. Consumers differ in their willingness to pay due to differences in income, age, education and/or other characteristics. The utility function for the consumption of product i is given by:

$$U_i = V + \lambda q_i - p_i \qquad (i=1,2) \qquad (4.1)$$

where p_i is the product price and V is a base level of utility. Product i will be consumed as long as $U_i \geq 0$ and $U_i > U_j$. It is assumed that V is large enough for $V \geq p_i \; \forall i=1,2$ to hold true so that the market is always served by at least one product.

It is assumed that the patentee's product is product 1 of quality q_1 and the entrant's product is product 2 of quality q_2. It is also assumed that the entrant enters with a better quality product, such that $q_2=q_1+\varepsilon$ where $\varepsilon \in (0, (\bar{q}-q_1)]$. The parameter ε represents the difference between the entrant's and the patentee's quality or equivalently the distance away from the patentee's product that the entrant locates in the quality space. The consumer who is indifferent between products 1 and 2 has a λ value denoted by $\hat{\lambda}$ and given by:

$$U_1 = U_2 \Rightarrow \hat{\lambda} = \frac{(p_2 - p_1)}{(q_2 - q_1)} = \frac{(p_2 - p_1)}{\varepsilon} \qquad (4.2)$$

The marginal consumer $\hat{\lambda}$ determines the market shares of products q_1 and q_2 as depicted in Figure 4.1.

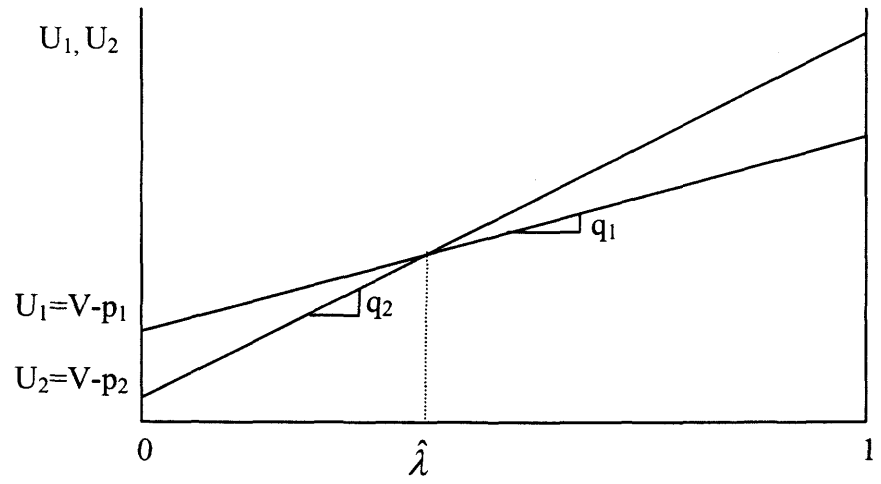


Figure 4.1 Market Shares of Product 1 of Quality q_1 and Product 2 of Quality q_2

Consumers with a λ value such that $\lambda \in [0, \hat{\lambda}]$ buy the patentee's product (the lower quality product) while consumers with a λ value such that $\lambda \in (\hat{\lambda}, 1]$ buy the entrant's product (the higher quality product). Given that every consumer buys only one unit of the product of his choice, the demand the patentee and the entrant face for their products are given by $y_1 = \hat{\lambda}$ and $y_2 = 1 - \hat{\lambda}$, respectively.

The same production cost structure is assumed for both the patentee and the entrant. The per unit production costs are assumed to be independent of the level of quality and are equal to zero. The production of product 2 of a given quality q_2 requires the incurring of fixed sunk costs (R&D costs), however. These fixed costs, denoted by $F_R(q)$, are an increasing function of quality: $F_R = \beta \frac{q^2}{2}$ where $\beta \geq \frac{4}{9}$.[2] The parameter β represents the effectiveness of the R&D process. Low β values represent high R&D

[2] See sub-section 4.3.1 for justification for the values of β.

effectiveness and imply low R&D costs while high β values represent low R&D effectiveness and imply high R&D costs.

The market conditions described above imply that the Finiteness Property introduced by Shaked and Sutton (1982) is satisfied. The Finiteness Property holds that a market where products are vertically differentiated, where the burden of quality improvements falls on fixed rather than variable costs and where the unit variable costs increase in quality slower relative to the willingness to pay for quality will be concentrated irrespective of its size and the level of fixed costs. Markets where the above conditions are present are natural oligopolies. The number of products in markets where the Finiteness Property holds is determined by the pattern of consumer tastes and income distribution (i.e., factors that differentiate consumers with respect to their willingness to pay for quality) (Shaked and Sutton 1982, 1983). In the present model, it is assumed that consumers' tastes and/or income distribution are such that the market can sustain only two products, that is, the market under consideration is a natural duopoly.

An important assumption in the model is that reverse engineering is possible and costless, which, in the absence of protection, enables the entrant to reproduce the patented product without incurring the R&D costs. Costless reverse engineering implies that $F_R(q_1) = 0$ for the entrant. Since $q_2=q_1+\varepsilon$ the R&D costs incurred by the entrant for the production of q_2 are a function of ε, namely: $F_R = \beta\frac{\varepsilon^2}{2}$. Given the above, it is increasingly costly for the entrant to locate away from the patentee (i.e., to produce the better quality product) in the quality product space. As will be shown below, the more costly it is for the entrant to produce product 2, the smaller is the degree of differentiation between her product and the patentee's product.

The patentee and the entrant compete in the one-dimensional product space presented in Figure 4.2 where quality is depicted on the vertical axes. In this product space point A represents product 1 of quality q_1. The breadth of protection claimed and granted to product 1 when it is patented is denoted by the variable b. The breadth of patent protection b takes values in the interval $b \in [0,\ \bar{q} - q_1]$. To normalize the model, it is assumed that $d = \bar{q} - q_1 = 1$ which implies that $b \in [0, 1]$ and $\varepsilon \in (0, 1]$. When b takes its minimum value of zero the protected area is just a point in the product space, point A. This case represents the minimum breadth of protection granted by the patent, namely zero breadth. With zero breadth of protection, the entrant can locate anywhere in the product space except at point A. Thus, zero breadth protects the patented product only against duplication.

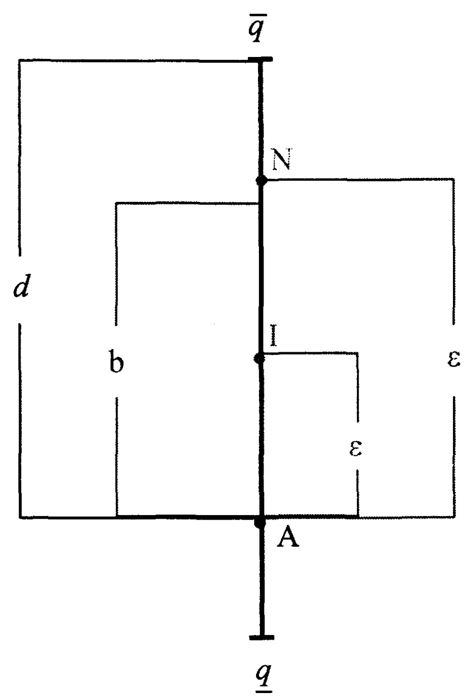


Figure 4.2 The Product Space and the Breadth of Patent Protection

narrow patents and invalidate broad ones (Waterson 1990, Cornish 1989, Merges and Nelson 1990). It is also assumed that when the maximum breadth is claimed (b=1), the patent will always be found to be invalid. These assumptions are captured by assuming that the probability that infringement is found, μ, is equal to $\mu = 1 - b$.[5]

The patent system being modeled is assumed to be that of the fencepost type, in which claims define an exact border of protection. Under the fencepost system, infringement will always be found when an entrant locates within the patentee's claims, unless the entrant proves that the patent is invalid (Cornish 1989).[6] In the fencepost system the probability that infringement is found does not depend on how close the entrant has located to the patentee. The implication of assuming a fencepost patent system is that the probability that infringement will be found (given that the entrant has located at $\varepsilon < b$ distance away from q_1) is equal to the probability that the validity of the patent will be upheld. Thus, the fencepost patent system implies that the events that the patent is found to be infringed and that the patent is found to be invalid can be treated as mutually exclusive and exhaustive.

A summary of the formal three stage strategic patent breadth determination game is presented diagrammatically in Figure 4.3. In Stage one, the patentee chooses patent breadth b. In Stage two, the entrant determines whether to enter in the patentee's market. If the entrant decides not to enter, the patentee makes monopoly profits (Π_m) in Stage three of the game and the entrant makes zero profits (see payoffs at A). If the entrant decides to enter she chooses where to locate in the quality space by choosing the

[5] To keep the model simple it is assumed that the patentee's and the entrant's trial costs do not affect the probability that infringement will be found (or equivalently that the validity of the patent will be upheld).
[6] In contrast, a signpost patent system implies that claims provide an indication of protection and the claims are interpreted using the doctrines of equivalents and reverse equivalents.

distance ε away from q_1. The choice of ε determines whether a trial occurs. The no trial outcome occurs if the entrant chooses $\varepsilon \geq b$. In this case, in Stage three of the game, the two competitors both produce their respective products and compete in the market by choosing prices. The payoffs for the patentee and the entrant are Π_P^{NI} and Π_E^{NI}, respectively (see payoffs at B). The trial outcome occurs if the entrant chooses $\varepsilon < b$. At trial, there is a probability μ that infringement is found and a probability $1-\mu$ that infringement is not found. If infringement is found during trial, the entrant is not allowed to market her product in Stage three of the game. In this case, in Stage three, the patentee operates as a monopolist while the entrant makes zero profits. If infringement is not found during the trial, then the patentee and the entrant compete by choosing prices. The entrant has no incentive to relocate within the quality space (i.e., the entrant has no incentive to move from point I in Figure 4.2) as she has already incurred the R&D costs which cannot be recovered. The payoffs for the patentee and the entrant when the entrant chooses $\varepsilon < b$ are $E(\Pi_P^I)$ and $E(\Pi_E^I)$, respectively (see payoffs at C).

After observing the quality of the patentee's product (q_1) and the breadth of the patent protection (b) granted to it, the entrant chooses whether or not to enter in the patentee's market. If the entrant decides to enter she has two choices: to locate inside ($\varepsilon < b$ – point I in Figure 4.2) or outside ($\varepsilon \geq b$ – point N in Figure 4.2) the patentee's claims. The first choice corresponds to a decision to infringe the patent, while the second corresponds to a decision not to infringe the patent. It is assumed that when the entrant locates at a distance $\varepsilon < b$ away from q_1, a trial always takes place, either because the patentee files an infringement lawsuit or because the entrant directly challenges the validity of the patent. It is further assumed that the filing of an infringement lawsuit is always met with a counterclaim by the accused infringer that the patent is invalid.[3] The costs incurred during the infringement trial/validity attack by the patentee and the entrant are denoted by C_P and C_E, respectively. These costs are assumed to be independent of the breadth of protection and of the entrant's location. The trial costs will only be incurred if $\varepsilon < b$ and they are assumed to be sunk – once made they cannot be recovered by either party.[4]

If the entrant locates within the patentee's claims the patent may not always be found to be valid. Indeed, the greater is the breadth of the patent, the smaller is the probability that the patent will be found to be valid or equivalently that infringement will be found. This assumption follows in part because the broader is the protection claimed, the harder it is to avoid obviousness, to differentiate the innovation from prior art (to show novelty) and to demonstrate that the innovation is enabling. As well, this assumption is in accordance with evidence from the literature that courts tend to uphold

[3] As described in chapter II, this is a standard defence adopted by accused infringers (Cornish 1989).
[4] With this assumption the possibility of the court awarding lawyers' fees to either party is excluded.

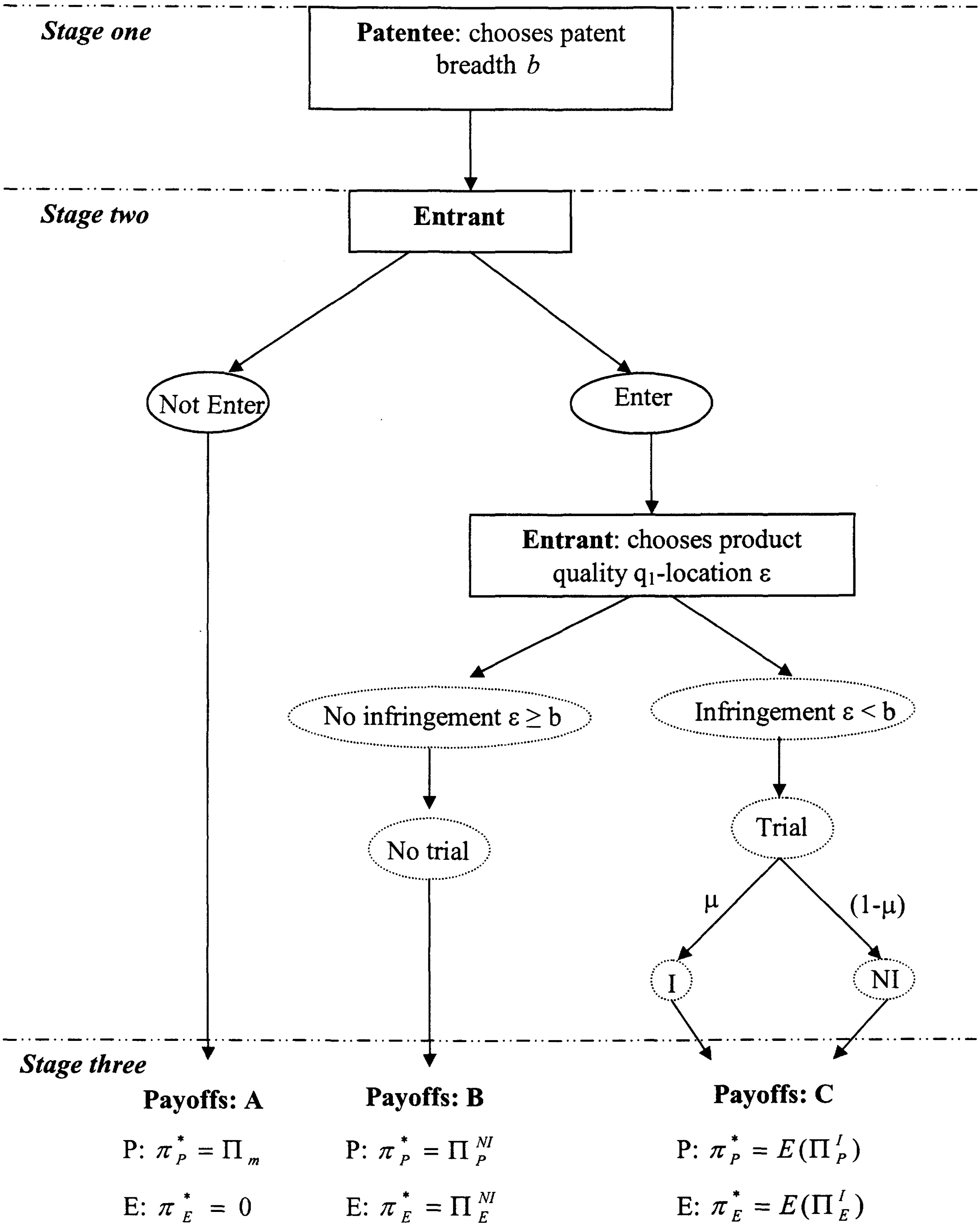


Figure 4.3 The Strategic Patent Breadth Game

4.3 The Analytical Solution of the Strategic Patent Breadth Game

The sub-game perfect Nash equilibrium of the sequential strategic patenting game is found using backwards induction. The prices that the patentee and the entrant charge for their products when they both operate in the market are determined first. The entrant's quality choice (her location on the quality product space) is derived next and the optimal breadth of patent protection for the patentee is determined last. The use of backwards induction eliminates multiple equilibria that do not represent credible threats and yields the only sub-game perfect Nash equilibrium of the game (Fudenberg and Tirole 1991).

The game is first solved under the assumption of *no patent protection* to determine where it is optimal for the entrant to locate when her choice is not constrained by the breadth of patent protection. This represents the entrant's most preferred location choice which is used as a benchmark for comparison with choices that are available to the entrant when her location choices are constrained by the breadth of patent protection.

4.3.1 No Patent Protection

- *The Entrant's Most Preferred Location Choice*

The entrant's most preferred location is found using backwards induction. Since it is assumed that there is no patent protection there are only two stages in this game. The equilibrium prices of the incumbent's and the entrant's products are determined first. The equilibrium prices determine the Bertrand profits for the two players. The entrant's quality choice, her location on the quality product space is determined last.

<u>The Pricing Equilibrium</u>. In this stage of the game the quality choice has been made and fixed costs have been sunk. As shown in sub-section 4.2.1 the demand for the

incumbent's and the entrant's product is given by $y_1 = \hat{\lambda}$ and $y_2 = 1 - \hat{\lambda}$, respectively. The incumbent and the entrant choose the prices for their products that maximize their profits, given respectively by:

$$\text{I:} \qquad \max_{p_1} \pi_1^B = p_1 y_1 = p_1 \frac{p_2 - p_1}{\varepsilon}$$
$$\text{E:} \qquad \max_{p_2} \pi_2^B = p_2 y_2 = p_2 (1 - \frac{p_2 - p_1}{\varepsilon}) \tag{4.3}$$

Optimization of the objective functions in (4.3) yields the following first order conditions (F.O.C.) for a maximum:

$$\frac{\partial \pi_1^B}{\partial p_1} = 0 \Rightarrow p_1^* = \frac{p_2}{2}$$
$$\frac{\partial \pi_2^B}{\partial p_2} = 0 \Rightarrow p_2^* = \frac{p_1 + \varepsilon}{2} \tag{4.4}$$

Simultaneously solving the equations in (4.4) yields the equilibrium prices, the quantities supplied and the profits in the final stage of the game, given by:

$$\text{I:} \qquad p_1^* = \frac{\varepsilon}{3},\ y_1^* = \frac{1}{3},\ \pi_1^B = \frac{\varepsilon}{9} \tag{4.5}$$
$$\text{E:} \qquad p_2^* = \frac{2\varepsilon}{3},\ y_2^* = \frac{2}{3},\ \pi_2^B = \frac{4\varepsilon}{9}$$

Since the entrant has the higher quality product, she charges the higher price. The entrant serves two thirds of the market, while the incumbent serves one third of the market. Profits are increasing in the distance ε between the incumbent's and the entrant's location. The greater is the difference in quality between the two products, the less intense is competition at the final stage of the game and the greater are the profits for both the incumbent and the entrant.[7]

<u>The Location Choice</u>. The entrant chooses the distance ε away from q_1 that will maximize her profits at this stage of the game. The objective function of the entrant is given by:

$$\text{E:} \qquad \max_{\varepsilon} \Pi_2^0 = \pi_2^B - F_R = \frac{4\varepsilon}{9} - \beta\frac{\varepsilon^2}{2} \tag{4.6}$$

Optimization of equation (4.6) yields the following F.O.C. for a maximum:

$$\frac{\partial \Pi_2^0}{\partial \varepsilon} = 0 \Rightarrow \frac{4}{9} - \beta\varepsilon_0 = 0 \Rightarrow \varepsilon_0 = \frac{4}{9\beta} \tag{4.7}$$

The second order conditions (S.O.C.) for a maximum are satisfied since:

$$\frac{\partial^2 \Pi_2^0}{\partial \varepsilon^2} < 0 \Rightarrow -\beta < 0\,, \qquad \forall \beta \geq \frac{4}{9} \tag{4.8}$$

The *most preferred* location choice for the entrant is given by ε_0 in equation (4.7) which holds for $\beta \geq \frac{4}{9}$ since ε∈(0, 1]. The quality of the entrant's product is given by:

$q_2 = q_1 + \varepsilon_0 = q_1 + \frac{4}{9\beta}$.

The profits for the incumbent and the entrant under the *no protection* outcome are obtained by substituting equation (4.7) into their respective profit functions. This substitution yields the following payoffs:

$$\text{I:} \qquad \Pi_1^0 = \frac{\varepsilon_0}{9} = \frac{4}{81\beta} \tag{4.9}$$

[7] This is a well-established result in the product differentiation literature in simultaneous games. When competitors first simultaneously choose their locations in the product space and then compete in prices they choose maximum differentiation to relax competition in the pricing stage that would curtail their profits (Lane 1980, Motta 1993, Shaked and Sutton 1982).

E: $$\Pi_2^0 = \frac{4\varepsilon_0}{9} - \beta\frac{(\varepsilon_0)^2}{2} = \frac{8}{81\beta}$$

Proposition 4.1 *Under no patent protection the less costly it is to produce the better quality product (i.e., the smaller is β), the further away from the incumbent the entrant locates and the greater are profits for both the incumbent and the entrant.*

Proof:

$$\frac{\partial\varepsilon_0}{\partial\beta} = -\frac{4}{9\beta^2} \leq 0 \ \forall\beta \geq \frac{4}{9};$$

$$\frac{\partial\Pi_1^0}{\partial\beta} = -\frac{4}{81\beta^2} \leq 0 \ \forall\beta \geq \frac{4}{9};$$

$$\frac{\partial\Pi_2^0}{\partial\beta} = -\frac{8}{81\beta^2} \leq 0 \ \forall\beta \geq \frac{4}{9}. \ \square$$

When the R&D costs are minimized – this happens when β takes its minimum value $(\beta = \frac{4}{9})$ – the entrant locates at the edge of the market (ε=1) maximizing differentiation between her product and the incumbent's product. The smaller are the R&D costs the greater is the distance ε away from the incumbent that the entrant locates and the greater are the profits for both the incumbent and the entrant. Thus, maximum possible product differentiation is desirable by both players.

4.3.2 Patent Protection $b \in [0,1]$

When the incumbent's product is protected by a patent, the entrant's location choices are constrained. Given the assumption of complete information, the patentee knows the entrant's cost structure, her trial costs and can anticipate the entrant's reaction to his choice of patent breadth. Since the entrant's location choice determines the level of the patentee's profits, the patentee chooses the breadth of protection that induces the desired behavior from the entrant.

The patentee knows that there is only one case in which the breadth of the patent does not influence the entrant's location decision. This happens when the entrant's cost structure is such that it is optimal for her to locate at the edge of the market (when $\beta = \frac{4}{9}$ then $\varepsilon_0 = 1$). In this case, irrespective of the breadth of the patent (ε is greater or equal to b for all $b \in [0, 1]$), the patent is never infringed and profits are maximized for both players. The patentee is free to choose any patent breadth, even the maximum breadth of protection, without triggering the trial outcome and having his patent invalidated.

For any value of $\beta > \frac{4}{9}$, the breadth of the patent may affect the entrant's location decision, which in turns affects the patentee's profits. If the patentee chooses $0 < b \leq \varepsilon_0$ it is always optimal for the entrant to enter and to locate at her most preferred location, namely $\varepsilon_0 = \frac{4}{9\beta}$ and no trial will occur. This outcome yields the same payoffs as the *no protection* outcome analyzed above.[8] However, if the patentee chooses $\varepsilon_0 < b \leq 1$ the entrant must first decide, depending on the value of patent breadth, whether to enter or not in the patentee's market. If she finds it profitable to enter, she

[8] The entrant always finds it optimal to enter since in this case her profits (given by equation (4.9)) are always positive.

must further decide whether to infringe or not the patentee's product.

A key element in the patentee's decision making is whether there is a value of patent breadth, $\hat{b} \in (\varepsilon_0, 1]$, that can deter market entry. If $\hat{b}$ exists, the patentee can choose this patent breadth and make monopoly profits. This outcome is illustrated in the payoffs at A in Stage three in Figure 4.3. It is assumed that the entrant decides not to enter when she is indifferent between entering and not entering the market. Thus, $\hat{b}$ is defined as the patent breadth that makes the expected profits that the entrant realizes when she infringes the patent ($E(\Pi_E^I)$) and the profits that she realizes when she does not infringe the patent (Π_E^{NI}) less than or equal to zero.

If there is no value of patent breadth that can deter entry in the patentee's market the patentee must find whether there is a value of patent breadth, denoted by $\tilde{b} \in (\varepsilon_0, 1]$, that makes the entrant indifferent between infringing and not infringing the patent. The variable $\tilde{b}$ thus makes the entrant's payoffs at B equal to the payoffs at C in Figure 4.3. Formally, if $\tilde{b}$ exists it should make Z_E given in equation (4.10) equal to zero.

$$Z_E = E(\Pi_E^I) - \Pi_E^{NI} \tag{4.10}$$

If $\hat{b}$ exists the patentee always chooses to deter entry. If $\hat{b}$ does not exist and $\tilde{b}$ exists, the patentee makes a decision of patent breadth by comparing his expected profits when the patent is infringed ($E(\Pi_P^I)$) and his profits when the patent is not infringed (Π_P^{NI}). The difference in the patentee's profits between those two scenarios is denoted by Z_P and is given by:

$$Z_P = E(\Pi_P^I) - \Pi_P^{NI} \tag{4.11}$$

If $Z_P > 0$ the patentee chooses a patent breadth that induces the entrant to infringe; a patent breadth that makes $Z_E > 0$. If $Z_P \leq 0$ the patentee chooses a patent breadth that results in non-infringement; a patent breadth that makes $Z_E \leq 0$. It is assumed that the entrant chooses not to infringe when she is indifferent between infringing and not infringing the patent.

Since the patentee's profits depend on the entrant's location on the quality product space, the patentee must first solve the entrant's location problem to be able to determine the breadth of protection claimed that maximizes his profits. In other words, the patentee must first determine the values of $\hat{b}$ and $\tilde{b}$, if they exist. Note that both $\hat{b}$ and $\tilde{b}$ are such that, $\hat{b}, \tilde{b} \in (\varepsilon_0, 1]$. As it has been discussed in sub-section 4.3.2 above, the entrant may find it optimal not to enter or to enter and infringe the patent if and only if $\varepsilon_0 < b \leq 1$; when $b \leq \varepsilon_0$ it is always optimal for the entrant to enter and to locate at her most preferred location ε_0, infringement does not occur and a trial does not take place. To determine the values $\hat{b}$ and $\tilde{b}$ the patentee needs to determine the entrant's expected profits when she infringes the patent and her profits when she does not infringe the patent when $\varepsilon_0 < b \leq 1$. The case where the entrant finds it optimal to infringe the patent is considered first.

4.3.2.1 The Entrant's Location Decision when $\varepsilon_0 < b \leq 1$

- *The Entrant's Expected Profits when she Infringes the Patent* ($\varepsilon < b$)

When the entrant infringes the patent the trial outcome is triggered. During trial it is determined whether infringement has occurred (or equivalently whether the patent is valid) or whether infringement has not occurred (or equivalently whether the patent is invalid).

The Pricing Equilibrium. If infringement is found during the trial, the entrant is not allowed to market her product and makes zero profits in the final stage of the game, while the patentee makes monopoly profits:

$$\begin{aligned} &\text{P:} \qquad \pi_1^I = \Pi_m \\ &\text{E:} \qquad \pi_2^I = 0 \end{aligned} \tag{4.12}$$

If infringement is not found, the entrant is allowed to remain in the market and to produce the quality of product that she has chosen. In this case, both the entrant and the patentee market their products and compete in prices in the third stage of the game. Their Bertrand profits are determined through the process described in the pricing equilibrium in sub-section 4.3.1 and are given by:

$$\begin{aligned} &\text{P:} \qquad \pi_1^B = \frac{\varepsilon_T}{9} \\ &\text{E:} \qquad \pi_2^B = \frac{4\varepsilon_T}{9} \end{aligned} \tag{4.13}$$

where ε_T is the entrant's optimal location choice under the *trial* outcome.

The Location Choice. The location of the entrant is determined through the optimization of her expected profits given by:

$$\text{E:} \qquad \max_{\varepsilon} \mathrm{E}(\Pi_E^I) = \mu \cdot \pi_2^I + (1-\mu) \cdot \pi_2^B - F_R - C_E = 0 + b \cdot \frac{4\varepsilon}{9} - \beta \frac{\varepsilon^2}{2} - C_E \tag{4.14}$$

Recall that the probability of the patent being found valid is $\mu = 1-b$. Optimization of the objective function in equation (4.12) yields the F.O.C. for a maximum:

$$\frac{\partial E(\Pi_E)^I}{\partial \varepsilon} = 0 \Rightarrow \frac{4}{9}b - \beta\varepsilon_T = 0 \Rightarrow \varepsilon_T = \frac{4}{9\beta}b \tag{4.15}$$

The S.O.C. for a maximum are satisfied $\forall\beta\geq\frac{4}{9}$:

$$\frac{\partial^2 E(\Pi_E)^I}{\partial \varepsilon^2} < 0 \Rightarrow -\beta < 0 \tag{4.16}$$

Equation (4.15) shows that when the entrant infringes the patent she finds it optimal to locate at a distance proportional to the breadth of the patent. Because there is uncertainty with respect to whether the entrant will be able to continue in the market, she 'underlocates'. In order to reduce the R&D costs, which are incurred with certainty, the entrant locates closer to the patentee than she would have done had infringement not been a possibility. Note that when the patentee chooses the maximum patent breadth (b=1) the entrant finds it optimal to locate at her most preferred location, $\varepsilon_T = \varepsilon_0$. This occurs because the entrant knows that she will win at trial with certainty since when $b = 1$ the patent is never found to be valid (i.e., $\mu = 1 - b = 0$).

The entrant's expected profits when she infringes the patent are obtained by substituting the entrant's optimal location under trial from equation (4.15) into her expected profit function. The substitution yields the following payoffs:

$$\text{E:} \qquad E(\Pi_E^I) = \frac{8}{81\beta}b^2 - C_E \tag{4.17}$$

Equation (4.17) shows that the greater is the breadth of the patent, the greater are the expected profits for the entrant under the trial outcome. This result occurs because the greater is the breadth of the patent, the greater is the probability that infringement will not be found (or equivalently that the patent will be invalidated) and thus, the greater is the probability that the entrant will be able to operate in the market. In addition, the greater are the trial costs that the entrant must incur the smaller are her expected profits when she infringes the patent. Figure 4.4 depicts the relationship between expected profits under infringement and the breadth of patent protection for different R&D effectiveness and trial cost values.

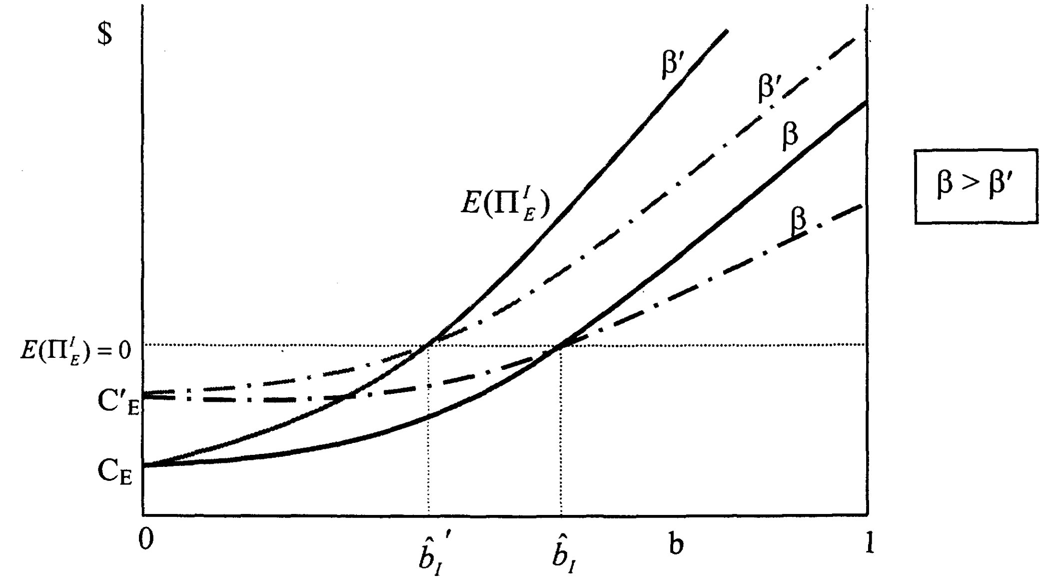


Figure 4.4 The Entrant's Expected Profits under Infringement

Equation (4.17) gives one set of conditions under which the entrant can be deterred from entering the market. The entrant will enter into the patentee's market *and* infringe the patent if and only if $E(\Pi_E^I) > 0$. Thus, the patentee can prevent entry (and subsequently patent infringement) by choosing a patent breadth that makes $E(\Pi_E^I) \leq 0$, i.e., by choosing a patent breadth that satisfies:

$$b \leq \sqrt{\frac{81\beta C_E}{8}} = \hat{b}_I \tag{4.18}$$

Thus, $\hat{b}_I$ denotes the breadth of patent protection that makes the entrant indifferent between entering the market *and* infringing the patent on the one hand and not entering on the other hand. Whether $\hat{b}_I \in (\varepsilon_0, 1]$ exists depends on the values of β and C_E.

Equation (4.18) shows that the greater are the costs of producing the better quality product (i.e., the greater is β) and the entrant's trial costs, the greater is the breadth of the patent that would prevent entry under infringement. These insights are depicted in Figure 4.4 and the intuition behind it them as follows. The greater are the costs of producing the entrant's product, the closer to the patentee the entrant is forced to locate and the smaller are the profits for the entrant. Similarly, the greater are the trial costs, the less profitable infringement becomes. Under these conditions, infringement is profitable only if patent breadth is relatively large. The greater is patent breadth the larger is the probability that the patent will be invalidated during trial and the greater is the probability that the entrant will be able to operate in the market.

- *The Entrant's Profits when she Does Not Infringe the Patent* ($\varepsilon \geq b$)

The Pricing Equilibrium. When the entrant does not infringe the patent both the patentee and the entrant market their products and compete in prices in the final stage of the game. Their profits are determined through the process described in the pricing equilibrium in sub-section 4.3.1 and are given by:

$$\begin{aligned} &\text{P:} \quad \pi_1^B = \frac{\varepsilon_n}{9} \\ &\text{E:} \quad \pi_2^B = \frac{4\varepsilon_n}{9} \end{aligned} \tag{4.19}$$

where ε_n is the optimal location choice when the entrant does not infringe the patent.

The Location Choice. The entrant's optimal location choice under no infringement is determined through the optimization of the profits given by:

E: $$\max_{\varepsilon} \Pi_E^{NI} = \pi_2^B - F_R = \frac{4}{9}\varepsilon - \frac{\beta}{2}\varepsilon^2 \qquad (4.20)$$

$$s.t. \qquad \varepsilon \geq b$$

The Lagrangean of the entrant's profit maximization problem is:

$$L = \frac{4}{9}\varepsilon - \frac{\beta}{2}\varepsilon^2 + \lambda(\varepsilon - b)$$

The Kuhn-Tucker conditions for a maximum are:

$$\frac{\partial L}{\partial \varepsilon} \leq 0 \Rightarrow \frac{4}{9} - \beta\varepsilon + \lambda \leq 0, \quad \varepsilon \geq 0, \text{ and } \varepsilon\frac{\partial L}{\partial \varepsilon} = 0$$

$$\frac{\partial L}{\partial \lambda} \geq 0 \Rightarrow \varepsilon - b \geq 0, \quad \lambda \geq 0, \text{ and } \lambda\frac{\partial L}{\partial \lambda} = 0$$

Since $\varepsilon \neq 0 \Rightarrow \frac{\partial L}{\partial \varepsilon} = 0 \Rightarrow \frac{4}{9} - \beta\varepsilon + \lambda = 0$.

<u>Case 1</u>. If $\lambda=0$ then $\varepsilon - b > 0$ and $\frac{\partial L}{\partial \varepsilon} = 0 \Rightarrow \varepsilon_n = \frac{4}{9\beta} = \varepsilon_0$. This solution is rejected since under this case $\varepsilon > b > \varepsilon_0$.

<u>Case 2</u>. If $\lambda>0$ then $\varepsilon_n = b$ and $\frac{\partial L}{\partial \varepsilon} = 0 \Rightarrow \lambda = \beta \cdot b - \frac{4}{9}$. The profits for the entrant under this case are given by substituting the solution $\varepsilon_n = b$ into the entrant's profit function. This substitution yields the following profits:

$$\Pi_E^{NI} = \frac{4}{9}b - \frac{\beta}{2}b^2 \qquad (4.21)$$

Equation (4.21) shows that the greater are the costs of producing the higher quality product (the greater is β) the smaller are the profits for the entrant when she decides not to infringe the patent and locates outside the patentee's claims. Figure 4.5 depicts the entrant's profits when she does not infringe the patent under different levels of R&D effectiveness.

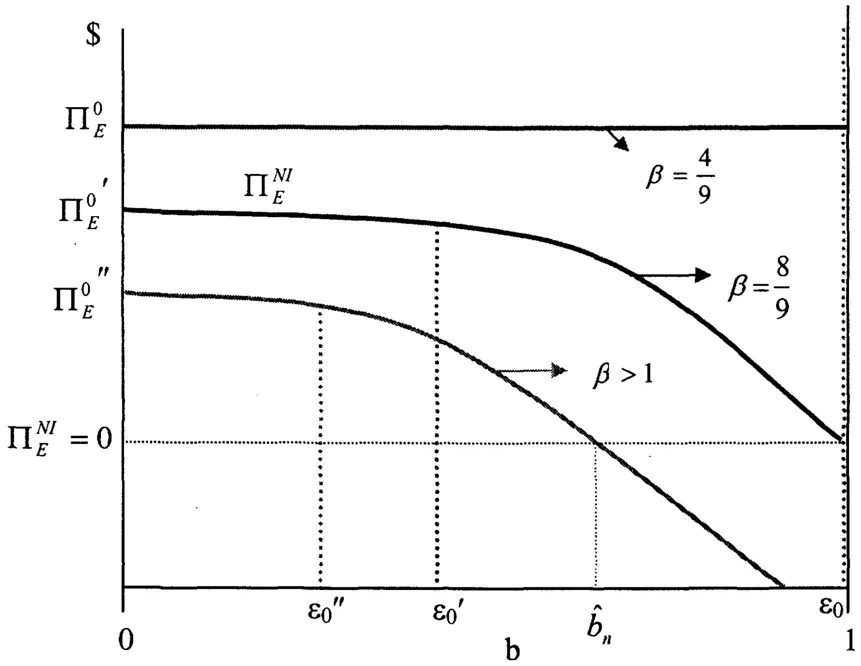


Figure 4.5 The Entrant's Profits under No Infringement

Figure 4.5 shows that the entrant's profits under no infringement are constant for breadth values less than or equal to the entrant's most preferred location ($b \le \varepsilon_0$) and they are declining for breadth values greater than the entrant's most preferred location ($\varepsilon_0 < b \le 1$). When patent breadth is less than or equal to the entrant's most preferred location, the entrant always finds it optimal to locate at her most preferred location (ε_0) and makes maximum profits (Π_E^0). However, when patent breadth is greater than the entrant's most preferred location it becomes increasingly costly for the entrant to locate outside the patentee's claims.

The entrant will choose to enter *and not* infringe the patent if and only if $\Pi_E^{NI} > 0$. The patent breadth that makes entry under no infringement non profitable for the entrant ($\Pi_E^{NI} \leq 0$) must satisfy the condition given by:

$$b \geq \frac{8}{9\beta} = \hat{b}_n \tag{4.22}$$

Thus, $\hat{b}_n$ denotes the breadth of patent protection that makes the entrant indifferent between entering the market *without infringing* the patent on the one hand and not entering on the other hand. Since $\hat{b}_n \in (\varepsilon_0, 1]$ equation (4.22) implies that $\hat{b}_n$ exists only for β values such that $\beta \geq \frac{8}{9}$.

Equation (4.22) shows that the greater are the R&D costs (the greater is β) that need to be incurred for the production of the entrant's product, the smaller is the patent breadth that prevents entry under no infringement. This result is depicted in Figure 4.5 and the intuition behind it is as follows. The entrant must locate outside the patentee's patent breadth for infringement not to occur. If patent breadth is large it may not be profitable for the entrant to locate outside the patentee's claims because it becomes more expensive to produce her product. The greater are the costs of producing the better quality product, the closer to the patentee the entrant is forced to locate and thus the smaller is the patent breadth that makes it unprofitable for the entrant to enter without infringing the patent.

Equations (4.18) and (4.22) give the conditions for non-entry under infringement and under no infringement, respectively. The breadth of patent protection that deters entry in the patentee's market, $\hat{b}$, if it exists, must simultaneously satisfy both conditions for non-entry under infringement and under no infringement. Thus, the entry deterrence condition is given by:

$$\hat{b}_n \leq \hat{b} \leq \hat{b}_I \Rightarrow \frac{8}{9\beta} \leq \hat{b} \leq \sqrt{\frac{81 C_E \beta}{8}} \tag{4.23}$$

Equation (4.23) shows that patent breadth $\hat{b}$ deters entry if and only if both the entrant's expected profits under infringement and her profits under no infringement are less or equal to zero; for $b = \hat{b}$ $E(\Pi_E^I) \leq 0 \wedge \Pi_E^{NI} \leq 0$.

Another important element in the patentee's decision making, besides the existence of the patent breadth $\hat{b}$ that can deter entry, is whether there is a patent breadth $\tilde{b} \in (\varepsilon_0, 1]$ that makes the entrant indifferent between infringing and not infringing the patent. If entry cannot be deterred (i.e., a $\hat{b}$ does not exist), before the entrant enters she must decide whether to infringe or not infringe the patent. As described in sub-section 4.3.2 patent breadth $\tilde{b}$, if it exists, makes the difference between the entrant's expected profits when she infringes the patent and her profits when she does not infringe the patent (denote by Z_E) equal to zero.

The determination of the entrant's expected profits under infringement and her profits under no infringement allow the patentee to determine the value of Z_E. Substitution of the expressions for the expected profits under infringement and the profits under no infringement given by equations (4.17) and (4.21), respectively, into the expression for Z_E, given by equation (4.10), yields:

$$Z_E = \left(\frac{8}{81\beta} + \frac{\beta}{2}\right)b^2 - \frac{4}{9}b - C_E \tag{4.24}$$

Equation (4.24) shows that Z_E is a function of the breadth of the patent (b), the entrant's cost structure (β) and the entrant's trial costs (C_E). The entrant's cost structure and the trial costs are exogenous to the game; these parameters are not affected by the decisions made by the patentee or the entrant. Patent breadth, however, is determined by the patentee. Thus, the breadth of patent protection claimed can determine whether the entrant will find it profitable to infringe or not infringe the patent.

Proposition 4.2 *When the entrant finds it optimal to enter the market (i.e., when entry cannot be deterred) then:*

(a) The greater is the breadth of patent protection the greater is the entrant's incentive to infringe the patent.

(b) The more costly it is to produce the better quality product the greater is the entrant's incentive to infringe the patent.

(c) The greater are the entrant's trial costs the smaller is the entrant's incentive to infringe the patent.

Proofs:

(α) $\frac{\partial Z_E}{\partial b} = (\frac{16}{81\beta} + \beta)b - \frac{4}{9} \geq 0 \ \forall \beta \geq \frac{4}{9} \wedge b \in (\varepsilon_0, 1]$

The greater is the breadth of patent protection the more costly it becomes for the entrant to locate outside the patentee's claims. In addition, the greater is the breadth of patent protection the greater is the probability that the patent will be invalidated and that the entrant will win at trial. Both the above outcomes increase the entrant's incentive to infringe the patent.

(b) $\frac{\partial Z_E}{\partial \beta} = (-\frac{8}{81\beta} + \frac{1}{2})b^2 \geq 0 \ \forall \ \beta \geq \frac{4}{9} \wedge b \in (\varepsilon_0, 1]$

The greater are the costs that need to be incurred for the production of the better quality product the less profitable it becomes for the entrant to locate outside the patentee's claims.

(c) $\frac{\partial Z_E}{\partial C_E} = -1 < 0$. □

The existence of a patent breadth that deters entry, $\hat{b}$, is closely linked to the existence of a patent breadth that makes the entrant indifferent between infringing and not infringing the patent, $\tilde{b}$. Figures 4.6 and 4.7 depict different scenarios with respect to the existence of $\hat{b}$ and $\tilde{b}$. Figure 4.6 depicts two cases under which entry cannot be deterred - a $\hat{b} \in (\varepsilon_0, 1]$ does not exist. Figure 4.7 depicts three cases under which entry can be deterred - a $\hat{b} \in (\varepsilon_0, 1]$ exists.

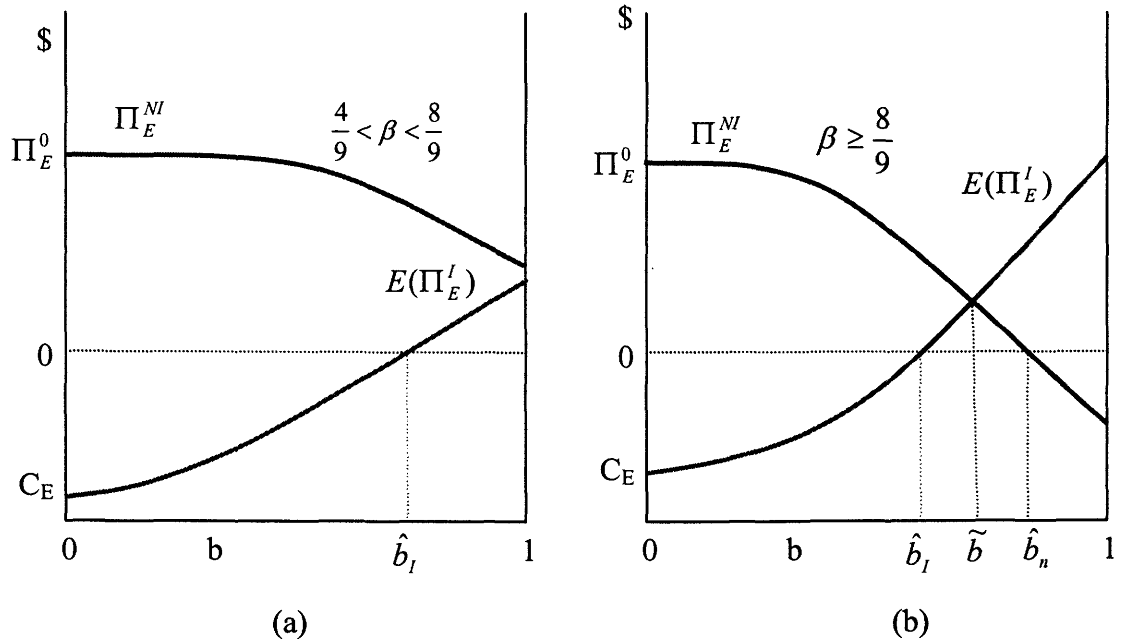


Figure 4.6 The Entrant's Profits under Infringement and No Infringement when Entry cannot be Deterred – a $\hat{b} \in (\varepsilon_0, 1]$ does not Exist

Panel (a) in Figure 4.6 represents the case where there is no patent breadth that can deter entry in the patentee's market and a $\widetilde{b} \in (\varepsilon_0, 1]$ does not exist. In this case non infringement is always an optimal strategy for the entrant as the curve depicting the entrant's profits under no infringement is above the curve depicting the entrant's expected profits under infringement for all patent breadth values (see Proposition 4.3 for a formal proof). Panel (b) in Figure 4.6 represents the case where there is a $\widetilde{b} \in (\varepsilon_0, 1]$, but $\widetilde{b}$ does not satisfy the entry deterrence condition, thus implying that entry cannot be deterred (see Proposition 4.6 for a formal proof). This result occurs because patent breadth $\widetilde{b}$ results in positive profits for the entrant irrespective of whether she infringes the patent or not. Neither $\hat{b}_I$ nor $\hat{b}_n$ can deter entry since none of them satisfies the entry deterrence condition.

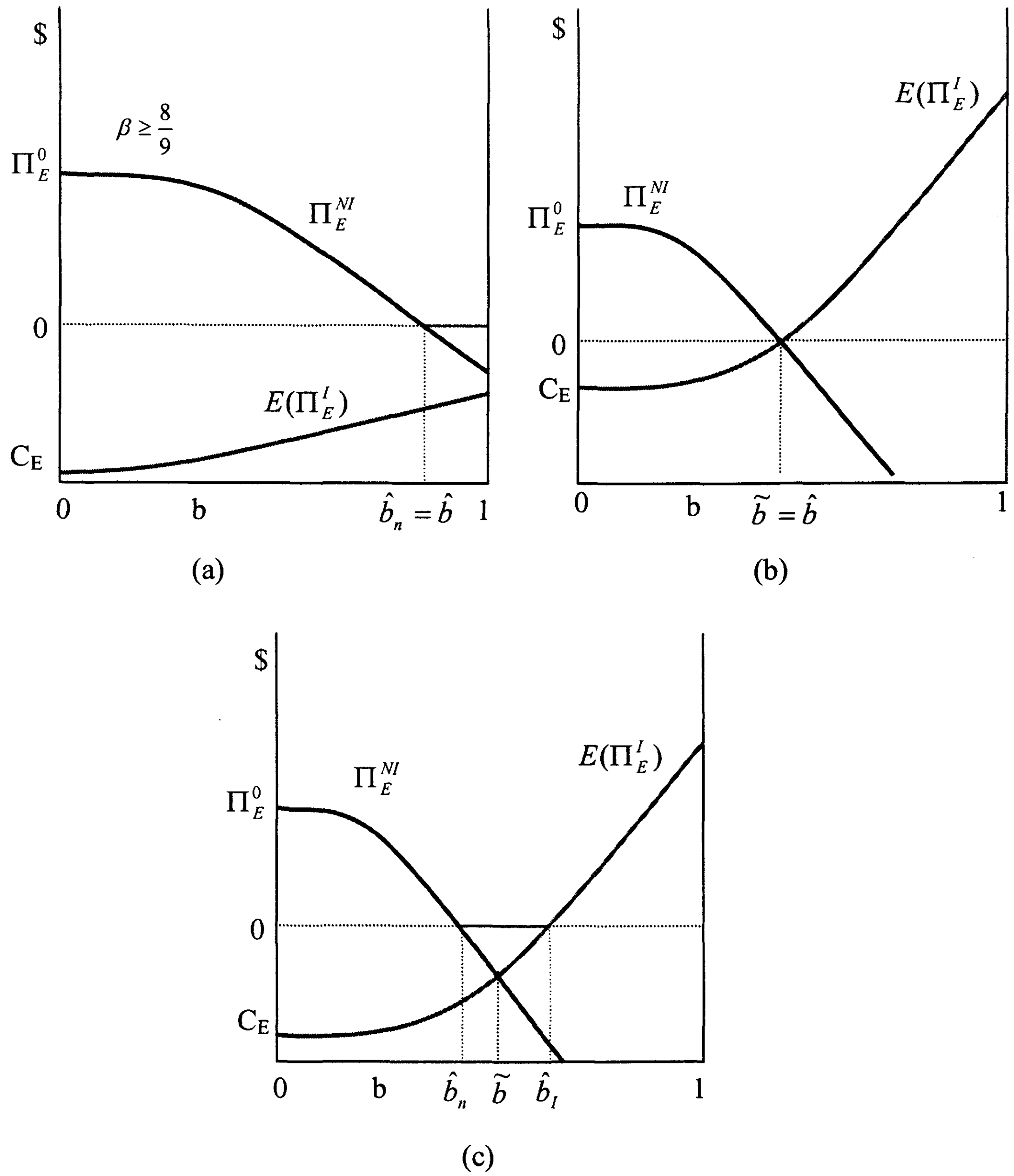


Figure 4.7 The Entrant's Profits under Infringement and No Infringement when Entry can be Deterred – a $\hat{b} \in (\varepsilon_0, 1]$ Exists

Panel (a) in Figure 4.7 represents the case where entry can be deterred and there is no $\tilde{b} \in (\varepsilon_0, 1]$. Patent breadth $\hat{b}_n$ deters market entry since it satisfies the entry deterrence condition. In fact, any value of patent breadth such that $b \in [\hat{b}_n, 1]$ can deter entry. Panel (b) in Figure 4.7 represents the case under which $\widetilde{b}$ is the only patent breadth that can deter entry. Finally, panel (c) in Figure 4.7 represents the case where there is a $\widetilde{b}$ ∈(ε_0, 1] and $\widetilde{b}$ satisfies the entry deterrence condition. In this case, there is a range of patent breadth values that can deter entry in the patentee's market. That is, either $\widetilde{b}, \hat{b}_I, \hat{b}_{II}$ or any $b \in [\hat{b}_{II}, \hat{b}_I]$ can deter entry since all the above patent breadth values satisfy the entry deterrence condition.

As it was mentioned above, patent breadth $\widetilde{b}$, if it exists, should make $Z_E = 0$. To determine whether a $\widetilde{b}$ exists $Z_E = (\frac{8}{81\beta} + \frac{\beta}{2})b^2 - \frac{4}{9}b - C_E = 0$ is solved for b. This solution yields the following two roots: $b_{1,2} = \dfrac{9(4\beta \pm \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2})}{16 + 81\beta^2}$.

The root $b_2 = \dfrac{9(4\beta - \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2})}{16 + 81\beta^2} \leq 0 \ \forall \ \beta \geq \frac{4}{9} \wedge C_E \geq 0$ and it is thus rejected since $\varepsilon_0 < \widetilde{b} \leq 1$. The root $b_1 = \dfrac{9(4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2})}{16 + 81\beta^2} \geq 0 \ \forall$ $\beta \geq \frac{4}{9} \wedge C_E \geq 0$ and it is accepted as a possible solution. Thus, if $\widetilde{b}$ exists it will be equal to b_1, i.e., $\widetilde{b} = b_1 = \dfrac{9(4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2})}{16 + 81\beta^2}$.

The patent breadth $\widetilde{b}$ that makes the entrant indifferent between infringing and not infringing the patent is a function of the entrant's cost structure (β) and her trial costs (C_E). Patent breadth $\widetilde{b}$ exists only if the values of β and C_E are such that $\varepsilon_0 < \widetilde{b} \leq 1$. It is easily verified that the condition $\widetilde{b} - \varepsilon_0 > 0$ is satisfied for all β and C_E values. That is,

$\tilde{b}-\varepsilon_0=\dfrac{9(4\beta+\sqrt{2}\sqrt{\beta}\sqrt{16C_E+8\beta+81C_E\beta^2})}{16+81\beta^2}-\dfrac{4}{9\beta}>0 \quad \forall \quad \beta\geq\dfrac{4}{9} \quad \wedge$ $C_E \geq 0$. The condition $\tilde{b}\leq 1$ is satisfied for certain combinations of β and C_E values. To determine the combinations of β and C_E values which satisfy the condition $\tilde{b}\leq 1$, the pairs of β and C_E values which satisfy the above constraint as an equality ($\tilde{b}=1$) are determined first. The solution of $\tilde{b}-1=0$ with respect to C_E yields: $C_E=\dfrac{16-72\beta+81\beta^2}{162\beta}$. The combination of β and C_E values for which $\tilde{b}-1=0$ is represented by the locus $\tilde{b}=1$ in Figure 4.8. The area to the right of the locus $\tilde{b}=1$ represents all combinations of β and C_E for which $\tilde{b}$ exists ($\tilde{b}<1$); this area includes the dotted and vertically hatched areas in Figure 4.8. This case is also depicted in panel (b) in Figure 4.6 and in panel (b) and (c) in Figure 4.7. The area to the left of the locus $\tilde{b}=1$ represents all combinations of β and C_E values for which $\tilde{b}$ does not exist ($\tilde{b}>1$); this area includes the non-shaded and the horizontally hatched areas in Figure 4.8. This case is also depicted in panel (a) in Figure 4.6 and panel (a) in Figure 4.7.

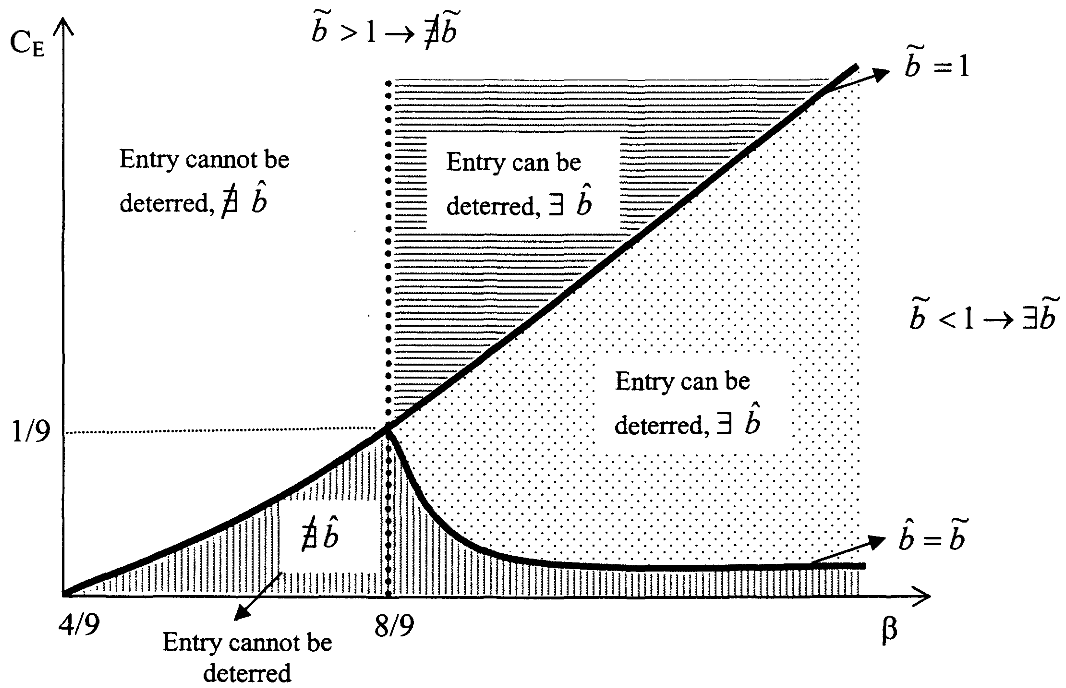


Figure 4.8 Combinations of C_E and β Values for which $\tilde{b}, \hat{b} \in (\varepsilon_0, 1]$ Exist

If $\varepsilon_0 < \tilde{b} \leq 1$ exists, it can deter entry if and only if it also satisfies the entry deterrence condition $\frac{8}{9\beta} \leq \tilde{b} \leq \sqrt{\frac{81 C_E \beta}{8}}$. The entry deterrence condition is satisfied for $\beta \geq \frac{8}{9}$ and for certain combinations of β and C_E values. To find the combinations of β and C_E values that satisfy the entry deterrence condition, the locus that satisfies the condition as an equality is determined first. The locus $\tilde{b} = \hat{b}$ in Figure 4.8 refers to the pairs of β and C_E values for which $\frac{8}{9\beta} = \tilde{b} = \sqrt{\frac{81 C_E \beta}{8}}$ holds true. Solution of the above condition with respect to C_E yields: $C_E = \frac{512}{6561\beta^3}$. This case is also depicted in panel (b) in Figure 4.7. All combinations of β and C_E values above the locus $\tilde{b} = \hat{b}$ and below the locus $\tilde{b} = 1$ – the dotted area in Figure 4.8 – satisfy the entry deterrence condition. This case is also depicted in panel (c) in Figure 4.7. The combinations of β and C_E values below the locus $\tilde{b} = \hat{b}$ and below the locus $\tilde{b} = 1$ – the vertically hatched area in Figure 4.8 – do not satisfy the entry deterrence condition. This case is also depicted in panel (b) in Figure 4.6.

The close relationship between the existence of a patent breadth $\hat{b} \in (\varepsilon_0, 1]$ that can deter entry and a patent breadth $\tilde{b} \in (\varepsilon_0, 1]$ that makes the entrant indifferent between infringing and not infringing the patent is further demonstrated in the propositions that follow.

Proposition 4.3 *If $\tilde{b} \in (\varepsilon_0, 1]$ does not exist it is never optimal for the entrant to infringe the patent.*

Proof:

At the entrant's most preferred location ε_0 non infringement is always more profitable than infringement for the entrant. That is, for $b = \varepsilon_0 = \frac{4}{9\beta}$ $Z_E = -C_E + \frac{128}{6561\beta^3} - \frac{8}{81\beta} < 0 \;\; \forall \;\; \beta \geq \frac{4}{9} \wedge C_E \geq 0$. The above conditions imply that if a $\tilde{b} \in (\varepsilon_0, 1]$ does not exist (i.e., there is no patent breadth that makes $Z_E=0$), then $Z_E<0$ $\forall b \in [0,1]$ which implies that $\Pi_E^{NI} > E(\Pi_E^I)$. This result is depicted in panel (a) in Figure 4.6 and in panel (a) in Figure 4.7. □

Proposition 4.4 *If $\tilde{b} \in (\varepsilon_0, 1]$ does not exist, the only patent breadth $\hat{b} \in (\varepsilon_0, 1]$ that can deter entry is the patent breadth that satisfies the non-entry condition under no infringement.*

Proof:

From Proposition 4.3 it is known that for $b=\varepsilon_0$ $Z_E<0$. If $\tilde{b}$ that makes $Z_E=0$ does not exist then $\forall b \in [0,1]$ $Z_E<0 \Rightarrow \Pi_E^{NI} > E(\Pi_E^I)$. If there is a patent breadth $\hat{b}_n$ that satisfies the non-entry condition under no infringement this implies that for $b=\hat{b}_n$ $\Pi_E^{NI} \leq 0$. Given that $\Pi_E^{NI} > E(\Pi_E^I)$, when $b=\hat{b}_n$ the entry deterrence condition is also satisfied. In this case, any $b \in [\hat{b}_n, 1]$ can deter entry. This case is depicted in panel (a) in Figure 4.7. □

Proposition 4.5 *If $\tilde{b} \in (\varepsilon_0, 1]$ exists:*

(a) The greater are the costs of producing the higher quality product, the smaller is the breadth of the patent that makes the entrant indifferent between infringing and not infringing the patent.

(b) The greater are the trial costs, the greater is the breadth of the patent that makes the entrant indifferent between infringing and not infringing the patent.

Proofs:

(α) $$\frac{\partial \tilde{b}}{\partial \beta} = \frac{9\left(4 + \dfrac{\sqrt{\beta}(8+162C_E\beta)}{\sqrt{2}\sqrt{16C_E+8\beta+81C_E\beta^2}} + \dfrac{\sqrt{16C_E+8\beta+81C_E\beta^2}}{\sqrt{2}\sqrt{\beta}}\right)}{16+81\beta^2} - \frac{1458\beta(4\beta+\sqrt{2}\sqrt{\beta}\sqrt{16C_E+8\beta+81C_E\beta^2})}{(16+81\beta^2)^2} \leq 0, \forall \beta \geq \frac{4}{9} \wedge C_E \geq 0$$

The more costly it is to produce the better quality product, the closer the entrant is forced to locate to the patentee and the smaller is the breadth of patent protection that makes it unprofitable for the entrant to locate above the patentee's patent breadth.

(b) $\dfrac{\partial \tilde{b}}{\partial C_E} = \dfrac{9\sqrt{\beta}}{\sqrt{2}\sqrt{16C_E + 8\beta + 81C_E\beta^2}} \geq 0 \;\; \forall \beta \geq \dfrac{4}{9} \wedge C_E \geq 0$

The greater are the trial costs the less appealing is infringement to the entrant. The entrant in this case will infringe the patent only if the breadth is so large that her cost structure does not allow her to locate outside the patentee's claims. □

Proposition 4.6 *If $\tilde{b} \in (\varepsilon_0, 1]$ exists and $\tilde{b}$ cannot deter entry (i.e., $\tilde{b}$ does not satisfy the entry deterrence condition), then there is no $\hat{b} \in (\varepsilon_0, 1]$ that can deter entry.*

Proof:

If $\tilde{b}$ exists, then for $b = \tilde{b}$ $Z_E = 0$. Since $\tilde{b}$ cannot deter entry it follows from equation (4.10) that at $b = \tilde{b}$ both $E(\Pi_E^I) > 0 \wedge \Pi_E^{NI} > 0$ must be satisfied. Assume that there is a $\hat{b} > \tilde{b}$ that can deter entry in the market. Then at $b = \hat{b}$ both $E(\Pi_E^I) < 0 \wedge \Pi_E^{NI} < 0$ should be satisfied. But $\dfrac{\partial E(\Pi_E^I)}{\partial b} \geq 0$ which, given that at $b = \tilde{b}$ $E(\Pi_E^I) > 0$, implies that $\forall \hat{b} > \tilde{b}$ $E(\Pi_E^I) > 0$. Thus, there is no patent breadth $\hat{b} > \tilde{b}$ that can deter entry. Now assume that there is a $\hat{b} < \tilde{b}$ that can deter entry in the market. Then at $b = \hat{b}$ both $E(\Pi_E^I) < 0 \wedge \Pi_E^{NI} < 0$ must be satisfied. But Π_E^{NI} is concave in b, $\dfrac{\partial \Pi_E^{NI}}{\partial b} \geq 0, \dfrac{\partial^2 \Pi_E^{NI}}{\partial b^2} \leq 0$ which, given that at $b = \tilde{b}$ $\Pi_E^{NI} > 0$, implies that $\forall \hat{b} < \tilde{b}$ at $b = \hat{b}$ $\Pi_E^{NI} > 0$. Thus, there is no patent breadth $\hat{b} < \tilde{b}$ that can deter entry in the market. This case is presented in Figure 4.6, panel (b). □

Proposition 4.7 *If* $\tilde{b} \in (\varepsilon_0, 1]$ *exists and it satisfies the entry deterrence condition as an equality then* $\tilde{b}$ *is the only breadth of patent protection that can deter entry.*

Proof:

The proof in this proposition is similar to the proof in Proposition 4.6. Since $\tilde{b}$ is the breadth of patent protection that makes Z_E=0, if $\tilde{b}$ makes $E(\Pi_E^I) = 0$ it should also make $\Pi_E^{NI} = 0$ (this follows from equation (4.10)). Since $\frac{\partial E(\Pi_E^I)}{\partial b} \geq 0$ $\forall$ $\hat{b} < \tilde{b}$ $E(\Pi_E^I) < 0$ and $\forall$ $\hat{b} > \tilde{b}$ $E(\Pi_E^I) > 0$. Also, since Π_E^{NI} is concave in b, $\forall$ $\hat{b} < \tilde{b}$ $\Pi_E^{NI} > 0$ and $\forall$ $\hat{b} > \tilde{b}$ $\Pi_E^{NI} < 0$. Thus, there is no $\hat{b} \neq \tilde{b}$ for which $E(\Pi_E^I) < 0$ $\wedge$ $\Pi_E^{NI} < 0$ which implies that there is no $\hat{b} \neq \tilde{b}$ that satisfies the entry deterrence condition. This case is depicted in Figure 4.7, panel (b). □

Proposition 4.8 *If* $\tilde{b} \in (\varepsilon_0, 1]$ *exists and it satisfies the entry deterrence condition as a strict inequality then there is a range of patent breadth values in the interval* $[\hat{b}_n, \hat{b}_I]$ *or in the interval* $[\hat{b}_n, 1]$ *that can deter entry.*

Proof:

If $\tilde{b}$ exists, then for $b = \tilde{b}$ Z_E=0. If $\tilde{b}$ can deter entry it follows from equation (4.10) that at $b = \tilde{b}$ both $E(\Pi_E^I) < 0 \wedge \Pi_E^{NI} < 0$ must be satisfied. Given that $\Pi_{E\,(b=0)}^{NI} > 0$, Π_E^{NI}

is concave in b and at $b = \tilde{b}$ $\Pi_E^{NI} < 0$, there is a breadth of patent protection $\hat{b}_n \in (0, \tilde{b})$ such that $\Pi^{NI}_{E\ (b=\hat{b}_n)} = 0$. Similarly given that $\frac{\partial E(\Pi_E^I)}{\partial b} \geq 0$ and at $b = \tilde{b}$ $E(\Pi_E^I) < 0$ there may exist a $\hat{b}_I \in (\tilde{b}, 1]$ such that $E(\Pi_E^I)_{(b=\hat{b}_I)} = 0$. This case is represented graphically in Figure 4.7, panel (c). If $\hat{b}_I \in (\varepsilon_0, 1]$ exists then any $b \in [\hat{b}_n, \hat{b}_I]$ can deter entry. If $\hat{b}_I \in (\varepsilon_0, 1]$ does not exist then any $b \in [\hat{b}_n, 1]$ can deter entry in the market. □

4.3.2.2 The Patentee's Strategic Patent Breadth Decision

In sub-section 4.3.2.1 it was shown that the existence of a patent breadth $\hat{b}$ that deters market entry and/or a patent breadth $\tilde{b}$ that makes the entrant indifferent between infringing and not infringing the patent depends on the entrant's R&D effectiveness (β) (i.e., her R&D cost structure) and her trial costs (C_E). The existence of $\hat{b}$ and $\tilde{b}$ determines the patentee's optimal patent breadth choice and the profits that can be realized. Different outcomes with respect to the patentee's patent breadth choice and profits emerge under different scenarios regarding the existence of $\hat{b}$ and $\tilde{b}$. These scenarios and the respective outcomes that emerge are presented in Figure 4.9.

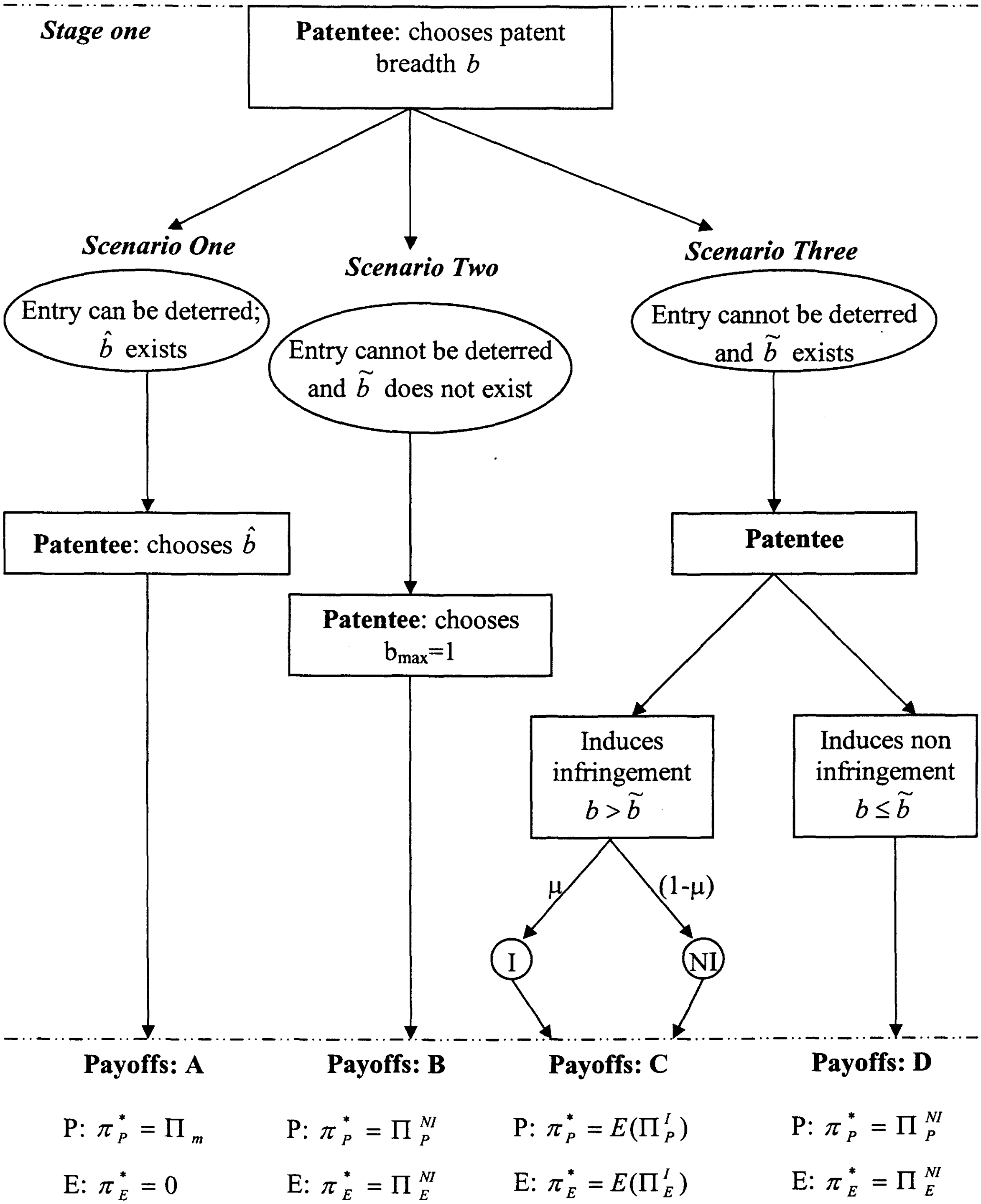


Figure 4.9 The Patentee's Strategic Patent Breadth Decision

- ***Scenario One*: There is a patent breadth $\hat{b}$ or a range of patent breadth values in the interval $[\hat{b}_n, \hat{b}_I]$ or in the interval $[\hat{b}_n, 1]$ that deter entry.**

Under this scenario, irrespective of whether $\tilde{b}$ exists or not, it is always optimal for the patentee to claim the breadth of patent protection $\hat{b}$ or any breadth values in the interval $[\hat{b}_n, \hat{b}_I]$ or in $[\hat{b}_n, 1]$ that deter entry. By claiming $\hat{b}$ the patentee makes monopoly profits Π_m.

- ***Scenario Two*: There is no patent breadth $\hat{b}$ that can deter entry and there is no patent breadth $\tilde{b}$ that makes the entrant indifferent between infringing and not infringing the patent.**

Under this scenario, as described in Proposition 4.3, the patent is never infringed. The patentee's profits under no infringement are $\Pi_P^{NI} = \pi_1^B = \frac{\varepsilon_n}{9}$, where $\varepsilon_n = b$ (see sub-section 4.3.2.1). The patentee chooses the breadth of patent protection that maximizes his objective function given by equation (4.25):

$$\text{P: } \max_b \Pi_P^{NI} = \frac{b}{9} \tag{4.25}$$

$$s.t. \qquad 0 \le b \le 1$$

Equation (4.25) shows that the patentee's profits under no infringement are increasing linearly in patent breadth. Given that patent breadth takes values in the interval $0 \le b \le 1$ the patentee's profits under no infringement are maximized for b=1. The patentee's profits when the patent is never infringed are depicted in Figure 4.10.

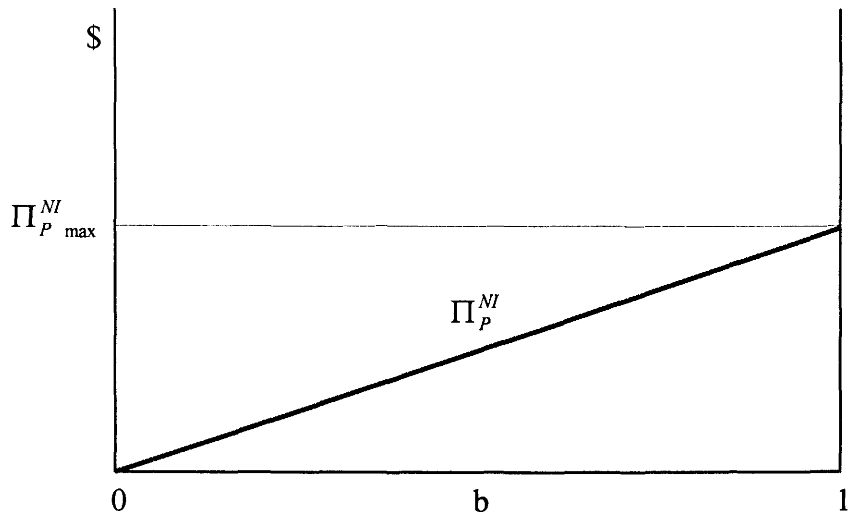


Figure 4.10 The Patentee's Expected Profits when the Patent is Never Infringed

In Figure 4.10 the patentee's profits are increasing linearly in patent breadth and are maximized for b=1. The above results show that when it is never optimal for the entrant to infringe the patent (i.e., when there is no $\hat{b}$ or $\tilde{b}$) it is always optimal for the patentee to claim the maximum breadth of patent protection, $b_{\max} = 1$.

The results in scenario two capture the standard assumption made in the patent breadth literature with respect to the patentee's patent breadth decision. The assumption in the patent literature is that the patentee claims the maximum breadth of patent protection (Merges and Nelson 1990, Gilbert and Shapiro 1990, Lanjouw and Schankerman 2001). The above result suggests claiming the maximum patent breadth is an optimal strategy for the patentee only if non infringement is an optimal strategy for the entrant.

- ***<u>Scenario Three</u>: There is a patent breadth $\tilde{b}$ that makes the entrant indifferent between infringing and not infringing the patent and $\tilde{b}$ cannot deter entry.***

Under this scenario, as it has been shown in Proposition 4.6, if $\tilde{b}$ cannot deter entry then there is no other breadth of patent protection $\hat{b} \in (\varepsilon_0,1]$ that can deter entry. In this case, the patentee has to determine whether it is more profitable to induce infringement by claiming a patent breadth $b > \tilde{b}$ or not to induce infringement by claiming a patent breadth $b \leq \tilde{b}$. The patentee under this scenario uses the value of $Z_P = E(\Pi_P^I) - \Pi_P^{NI}$ to determine the optimal patent breadth. If $Z_P > 0$ the patentee chooses a $b > \tilde{b}$ that induces the entrant to infringe the patent. If $Z_P \leq 0$ the patentee chooses a $b \leq \tilde{b}$ that induces non infringement. The optimal patent breadth value is determined through the solution of the patentee's maximization of expected profits under infringement and under no infringement.

- *The Patentee's Expected Profits when he Induces Infringement ($b > \tilde{b}$)*

When the patentee claims $b > \tilde{b}$ he knows that the entrant's optimal strategy is to infringe the patent. The patentee makes monopoly profits with probability $\mu = 1 - b$ if his patent is found valid during trial (or equivalently if infringement is found) and duopoly profits with probability $1 - \mu = b$ if his patent is revoked (or equivalently if infringement is not found). The patentee's duopoly profits are given by $\pi_1^B = \frac{\varepsilon_T}{9}$ where $\varepsilon_T = \frac{4}{9\beta} b$ is the entrant's optimal location when she infringes the patent (see sub-section 4.3.2.1). The patentee also incurs trial costs denoted by C_P which are independent of the breadth of patent protection claimed.

The patentee chooses the breadth of patent protection that maximizes his expected profits under infringement. The patentee's objective function is given by:

$$\text{P:} \qquad \max_{b} E(\Pi_P^I) = \mu\Pi_m + (1-\mu)\pi_1^B - C_P \tag{4.26}$$

$$s.t. \qquad \widetilde{b} + e \leq b \leq 1 \text{ where } e \to 0$$

The Lagrangean of the patentee's profit maximization problem is given by:

$$L = (1-b)\Pi_m + \frac{4b^2}{81\beta} - C_P + \lambda_1(1-b) + \lambda_2(b - \widetilde{b} - e)$$

The Kuhn-Tucker conditions for a maximum are:

$$\frac{\partial L}{\partial b} \leq 0 \Rightarrow -\Pi_m + \frac{8b}{81\beta} - \lambda_1 + \lambda_2 \leq 0,\ b \geq 0 \text{ and } b\frac{\partial L}{\partial b} = 0$$

$$\frac{\partial L}{\partial \lambda_1} \geq 0 \Rightarrow 1 - b \geq 0,\ \lambda_1 \geq 0 \text{ and } \lambda_1 \frac{\partial L}{\partial \lambda_1} = 0$$

$$\frac{\partial L}{\partial \lambda_2} \geq 0 \Rightarrow b - \widetilde{b} - e \geq 0,\ \lambda_2 \geq 0 \text{ and } \lambda_2 \frac{\partial L}{\partial \lambda_2} = 0$$

Since $\widetilde{b} + e \leq b \leq 1 \Rightarrow b \neq 0 \Rightarrow \frac{\partial L}{\partial b} = 0 \Rightarrow -\Pi_m + \frac{8b}{81\beta} - \lambda_1 + \lambda_2 = 0$

<u>Case 1.</u> If $\lambda_1 = \lambda_2 = 0 \Rightarrow \widetilde{b} + e < b < 1$ and from $\frac{\partial L}{\partial b} = 0 \Rightarrow b_I = \frac{81\beta\Pi_m}{8}$

The S.O.C. for a maximum are not satisfied, $\frac{\partial^2 L}{\partial b^2} = \frac{8}{81\beta} > 0$ which implies that b_I is a minimum not a maximum and b_I is thus rejected as a solution. The above conditions indicate that there is a corner solution to the expected profit maximization problem. Thus, either b=1 or $b = \widetilde{b} + e$ is the breadth of patent protection that maximizes the patentee's expected profits under infringement.

<u>Case 2.</u> If $\lambda_1 > 0$ then b=1 and $\lambda_2 = 0$. In this case, $\lambda_1 = \frac{8}{81\beta} - \Pi_m$ and the patentee's

expected profits are:

$$E(\Pi_P^I)_{b=1} = \frac{4}{81\beta} - C_P \tag{4.27}$$

Case 3. If $\lambda_2 < 0$ then $b = \widetilde{b} + e$ and $\lambda_1 = 0$. In this case, $\lambda_2 = \Pi_m - \frac{8(\widetilde{b}+e)}{81\beta}$ and the patentee's expected profits are:

$$E(\Pi_P^I)_{b=\widetilde{b}+e} = (1-\widetilde{b}-e)\Pi_m + \frac{4}{81\beta}(\widetilde{b}+e)^2 - C_P$$

$$\lim_{e\to 0} E(\Pi_P^I)_{b=\widetilde{b}+e} = \Pi_m - \widetilde{b}\Pi_m + \frac{4}{81\beta}\widetilde{b}^2 - C_P \tag{4.28}$$

where $\widetilde{b} = \dfrac{9(4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2})}{16+81\beta^2}$

Comparison of the patentee's expected profits when b=1 given by equation (4.27) to the patentee's expected profits when $b = \widetilde{b} + e$ given by equation (4.28) yields the following results. For monopoly profit values $\Pi_m \geq 0.089$, $E(\Pi_P^I)_{b=1} \leq E(\Pi_P^I)_{b=\widetilde{b}+e}$. For monopoly profit values $\Pi_m < 0.089$, $E(\Pi_P^I)_{b=1} > E(\Pi_P^I)_{b=\widetilde{b}+e}$. Note that, under scenario three, all values of the entrant's R&D effectiveness (β) and trial costs (C_E) are such that both $C_E \leq \frac{16-72\beta+81\beta^2}{162\beta}$ and $C_E \leq \frac{512}{6561\beta^3}$ are satisfied (i.e., β and C_E values in the vertically hatched area in Figure 4.8).

The above results show that the smaller are the monopoly profits that the patentee makes when his patent is found valid at trial, the greater is the patentee's incentive to claim the maximum breadth of protection and have his patent revoked. This happens because under infringement the entrant's location is proportional to the breadth of the patent (i.e., $\varepsilon_T = \frac{4}{9\beta}b$) so the greater is patent breadth the further away from the patentee the entrant locates and the greater are the profits at the last stage of the game for both players. In other words, in this case, the effect of the loss of monopoly profits due to the large patent breadth is smaller than the effect of the increased profits brought by the increased level of differentiation between the two products. However, when monopoly profits are large the patentee does not want to risk having his patent revoked by claiming the maximum breadth of patent protection and he claims $b = \widetilde{b} + e$ instead. In this case, the effect of the decrease in expected profits due to the decrease in product differentiation is smaller than the effect of the increase in expected profits due to the increased probability that infringement will be found at trial and the patentee will realize monopoly profits.

Given that the level of monopoly profits is unknown, two cases have to be considered. Under the first case $E(\Pi_P^I)_{b=1} > E(\Pi_P^I)_{b=\widetilde{b}+e}$, while under the second case $E(\Pi_P^I)_{b=1} \leq E(\Pi_P^I)_{b=\widetilde{b}+e}$. The two graphs in Figure 4.11 depict the above two cases when $b_I \in (\widetilde{b}+e,\, 1)$ (panel (a)) and when $b_I \notin (\widetilde{b}+e,\, 1)$ (panel (b)).

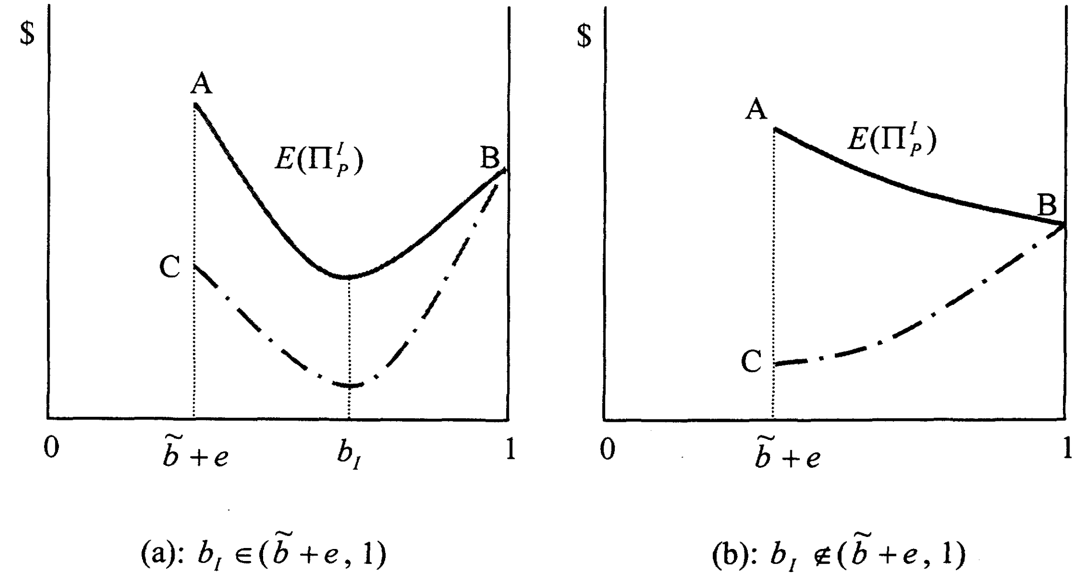


Figure 4.11 The Patentee's Expected Profits when Infringement is Induced

As shown in Figure 4.11 when the patentee's expected profits under infringement are represented by the curve AB the patentee maximizes his profits by choosing the patent breadth $b = \widetilde{b} + e$. When the curve CB reflects the patentee's expected profits under infringement then the patentee maximizes his profits by choosing the maximum breadth of patent protection $b_{\max} = 1$.

- *The Patentee's Profits when he Induces Non Infringement ($b \leq \widetilde{b}$)*

When the patentee claims $b \leq \widetilde{b}$ he knows that the entrant's optimal strategy is to not infringe the patent. The patentee chooses the breadth of patent protection that maximizes his profits under no infringement given by:

$$\text{P:} \qquad \max_b \Pi_P^{NI} = \pi_1^B = \frac{\varepsilon_n}{9} = \frac{b}{9} \tag{4.29}$$

$$s.t. \quad 0 \leq b \leq \widetilde{b}$$

Since the patentee's profits under no infringement are increasing linearly in patent breadth the breadth of patent protection that maximizes equation (4.29) is $b_n = \widetilde{b}$. Substituting $b_n = \widetilde{b}$ in equation (4.29) yields the patentee's maximum profits under no infringement:

$$\Pi_P^{NI} = \frac{\widetilde{b}}{9} = \frac{4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2}}{16 + 81\beta^2} \tag{4.30}$$

Having determined the optimal patent breadth and the level of profits under infringement and under no infringement the patentee can determine the value of Z_P. Two cases must be considered depending on whether $b = 1$ or $b = \widetilde{b} + e$ is the optimal patent breadth under infringement.

I. $E(\Pi_P^I)_{b=1} > E(\Pi_P^I)_{b=\widetilde{b}+e}$

Under this case Z_P is redefined as $Z_P^1 = E(\Pi_P^I)_{b=1} - \Pi_P^{NI}$.

Proposition 4.9 *When $\widetilde{b} \in (\varepsilon_0, 1]$ exists and it cannot deter entry, claiming the maximum breadth of patent protection ($b_{\max} = 1$) is never an optimal strategy for the patentee unless $\widetilde{b} = 1$.*

Proof:

$$Z_P^1 = E(\Pi_P^I)_{b=1} - \Pi_P^{NI} = \frac{4}{81\beta} - \frac{4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2}}{16 + 81\beta^2} - C_P < 0 \; \forall \beta \geq \frac{4}{9} \wedge$$

$C_P \geq 0 \wedge C_E \geq 0$. □

Since $Z_P^1 < 0$ the optimal strategy for the patentee under scenario three when $E(\Pi_P^I)_{b=1} > E(\Pi_P^I)_{b=\tilde{b}+e}$ is to claim patent breadth $b = \tilde{b}$ which does not induce infringement.

II. $E(\Pi_P^I)_{b=1} \leq E(\Pi_P^I)_{b=\tilde{b}+e}$

Under this case Z_P is redefined as $Z_P^2 = \lim_{e \to 0} E(\Pi_P^I)_{b=\tilde{b}+e} - \Pi_P^{NI}$.

Substituting equations (4.28) and (4.30) into Z_P^2 yields:

$$Z_P^2 = \Pi_m - \tilde{b}\Pi_m + \frac{4}{81\beta}\tilde{b}^2 - C_P - \frac{\tilde{b}}{9}$$

where $\tilde{b} = \dfrac{9(4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2})}{16 + 81\beta^2}$

The value of Z_P^2 cannot be determined without knowledge of the values of the parameters Π_m, β, C_P and C_E. When $\tilde{b} \in (\varepsilon_0, 1]$ exists and it cannot deter entry the optimal breadth of patent protection is either ($b = \tilde{b}$) or ($b = \tilde{b} + e$) depending on the relative values of the parameters Π_m, β, C_P and C_E.

Proposition 4.10 *When the patentee cannot deter entry (a $\hat{b}$ does not exist) and there exists a patent breadth $\tilde{b}$ that makes the entrant indifferent between infringing and not infringing the patent then:*

(a) The greater are the patentee's monopoly profits (Π_m) the greater is the patentee's incentive to induce infringement.

(b) The greater are the patentee's trial costs (C_P) the smaller is the patentee's

incentive to induce infringement.

(c) The greater are the entrant's costs of producing the better quality product the greater is the patentee's incentive to induce infringement given that the patentee's monopoly profits are different than zero.

Proofs:

(α) $\frac{\partial Z_P^2}{\partial \Pi_m} = 1 - \frac{9(4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2}}{16 + 81\beta^2} \geq 0$ ∀β, C_E such that,

$C_E \leq \frac{16 - 72\beta + 81\beta^2}{162\beta} \wedge C_E \leq \frac{512}{6561\beta^3}$.

Since under this scenario the patentee cannot deter entry, the only case that he can earn monopoly profits is if his patent is infringed and he wins at trial. Thus, the greater are the monopoly profits that he anticipates to make the greater is his incentive to claim a patent breadth that will induce infringement.

(b) $\frac{\partial Z_P^2}{\partial C_P} = -1 < 0$

(c) $\frac{\partial Z_P^2}{\partial \beta} = -A + 162\beta B + 8AB(16 + 81\beta^2) - \frac{1296B}{16 + 81\beta^2} - \frac{4B^2}{\beta}(16 + 81\beta^2)^2 +$
$\Pi_m(-9B(16 + 81\beta^2) + 1458\beta B) \geq 0$

∀β, C_E such that, $C_E \leq \frac{16 - 72\beta + 81\beta^2}{162\beta} \wedge C_E \leq \frac{512}{6561\beta^3}$ and $\Pi_m > 0$.

Where $A = \frac{4 + \frac{\sqrt{\beta}(8 + 162C_E\beta)}{\sqrt{2}\sqrt{16C_E + 8\beta + 81C_E\beta^2}} + \frac{\sqrt{16C_E + 8\beta + 81C_E\beta^2}}{\sqrt{2}\sqrt{\beta}}}{16 + 81\beta^2}$ and

$B = \frac{4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2}}{(16 + 81\beta^2)^2}$.

The intuition behind this result is as follows. The greater are the entrant's R&D costs, the closer the entrant is forced to locate to the patentee. In this case, the patentee has a greater incentive to induce infringement because the closer to the patentee the entrant is forced to locate, the smaller need be the patent breadth that will induce the entrant to infringe and thus, the smaller is the probability that the patent will be invalidated at trial. The effect that C_E has on the patentee's incentive to infringe the patent is inconclusive, it depends on the values of β and Π_m. □

Figure 4.12 depicts the patentee's expected profits under infringement and his profits under no infringement as a function of patent breadth.

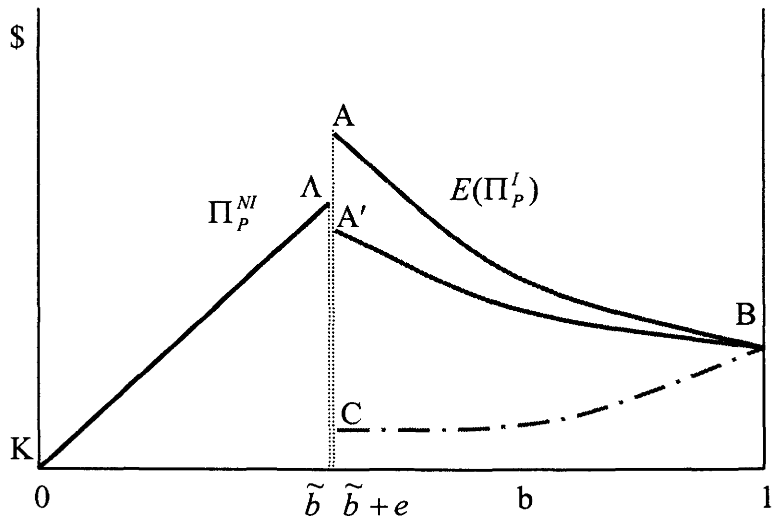


Figure 4.12 The Patentee's Expected Profits under Infringement and his Profits under No Infringement under Scenario Three

In Figure 4.12 line KΛ represents the profits that the patentee makes when his patent is not infringed. The curves AB and A'B refer to the expected profits that the patentee makes when his patent is infringed and $E(\Pi_P^I)_{b=1} \le E(\Pi_P^I)_{b=\tilde{b}+e}$. The curve CB refers to the expected profits that the patentee makes when his patent is infringed and $E(\Pi_P^I)_{b=1} > E(\Pi_P^I)_{b=\tilde{b}+e}$. When the patentee's expected profits under infringement are depicted by the curve A'B or the curve CB, the profits for the patentee are maximized at point Λ where the breadth of patent protection is $\tilde{b}$. When the patentee's expected profits under infringement are depicted by the curve AB profits for the patentee are maximized at point A where the breadth of patent protection is $\tilde{b}+e$.

To summarize the findings of sub-section 4.3.2.2, the patentee's choice of the optimal patent breadth depends on the entrant's R&D cost structure (β) and her trial costs (C_E). When the combination of β and C_E values is such that both conditions $\beta \ge \frac{8}{9}$ and $C_E \ge \frac{512}{6561\beta^3}$ are satisfied (β and C_E values are in the dotted and horizontally hatched areas in Figure 4.8), then there exists at least one patent breadth $\hat{b}$ that can deter entry in the market. In this case, the patentee always chooses patent breadth $\hat{b}$ and maximizes his profits (Π_m) operating as a monopolist.

When the combination of β and C_E values is such that both conditions $\frac{4}{9} \le \beta < \frac{8}{9}$ and $C_E > \frac{16-72\beta+81\beta^2}{162\beta}$ are satisfied (β and C_E values are in the non-shaded area in Figure 4.8), then there is no patent breadth that can deter entry or that can make the entrant indifferent between infringing and not infringing the patent. The patentee thus finds it optimal to claim the maximum patent breadth (b_{max}=1), since in this case the patent is never infringed and a trial never occurs.

Finally, when the combination of β and C_E values is such that both conditions $C_E \le \frac{16-72\beta+81\beta^2}{162\beta}$ and $C_E < \frac{512}{6561\beta^3}$ are satisfied (β and C_E values are in the vertically hatched area in Figure 4.8), then it is optimal for the patentee to choose either patent breadth $\tilde{b}$ and not induce infringement or $\tilde{b}+e$ and induce infringement. The choice of the optimal patent breadth in this case depends on the patentee's trial costs (C_P), the monopoly profits that the patentee will make if his patent is found valid at trial (Π_m) and the entrant's R&D cost structure (β) and trial costs (C_E).

4.4 Concluding Remarks

In this chapter a simple game theoretic model is used to describe the patenting behavior of an innovator who, having invented a drastic product innovation and having decided to seek patent protection, determines the breadth of patent protection that maximizes the appropriability of the rents from his innovation. To determine the optimal breadth of patent protection claimed, the patentee acts strategically, choosing the breadth of protection that induces the desired behavior by the entrant. The patentee is foresighted and anticipates that he may have to incur costs to enforce his patent rights. The model suggests that the strategic patent breadth, that is, the breadth of patent protection that maximizes the innovators ability to appropriate innovation rents, depends on the entrant's R&D cost structure and her trial costs.

Contrary to what it is traditionally assumed, the results show that it is not always optimal for the patentee to claim the maximum patent breadth possible. In fact, only for certain values of the entrant's R&D effectiveness and trial costs it is optimal for the patentee to claim the maximum breadth of patent protection. The patentee claims maximum patent protection when he cannot deter entry and the entrant always finds it optimal to not infringe the patent. The maximum breadth of protection is also claimed when the only patent breadth that deters entry $\hat{b} = \tilde{b}$ is equal to one which occurs for a specific combination of R&D and trial costs (i.e., for $\beta = \frac{8}{9}$ and $C_E = \frac{1}{9}$).

The results hold under the assumption of a fencepost patent system, which implies that the events that the patent is infringed and the patent is invalid can be treated as mutually exclusive and exhaustive. In addition, it has been assumed that the market can only support two products, and that the R&D process is deterministic. Relaxing the above assumptions is the focus of future research.

CHAPTER V

STRATEGIC PATENT BREADTH FOR DRASTIC PROCESS INNOVATIONS

5.1 Introduction

In chapter IV the patent breadth that maximizes the innovator's ability to appropriate innovation rents was determined for drastic product innovations. This chapter examines the strategic patenting behavior of an innovator who has invented a patentable process innovation and uses patent protection to maximize the appropriability of this innovation. Process innovations are new technologies that refer to the way or method of generating a product (e.g., an example of a process innovation could be a new process that produces a purified version of existing proteins). The product itself does not have to be new or patentable for the process to be patentable (e.g., purified versions of existing proteins are not patentable in some countries). Usually new processes are either cost reducing and/or they improve the efficiency of producing a product (e.g., a process that generates a purer form of a protein).

The distinction between product and process innovations is an important one as the nature of competition in a market depends on whether the patent protects a product or a process (Eswaran and Gallini 1996). Product patents usually grant greater protection than process patents (Merges and Nelson 1990). When the innovator holds a product patent, rivals cannot produce the same product even if a different or a more efficient process is used for its production.[1] However, when the innovator holds a process patent for producing a given product, but not a patent on the product itself, rival firms with non-infringing processes can produce the same product or different versions of it.

An example used in the literature to underline the significance of the distinction between product and process patents for the level of competition in the market is the t-PA (Tissue Plaminogen Activator) drug. This is a clot dissolving heart attack drug, which is the first commercially successful product derived using recombinant DNA technology (Eswaran and Gallini 1996, Merges and Nelson 1990). This drug was introduced by Genentech, which received a product and a process patent in the US but only a process patent in the United Kingdom (in the United Kingdom the product was found to be non-patentable). In the United Kingdom, fifteen other firms that were working on the generation of the drug were free to purse non-infringing processes for generating the non-patentable product. In the US, however, work by other firms on the drug was blocked, mainly because the company held a product patent (Eswaran and Gallini 1996).

The process innovations that are examined in this chapter are assumed to be drastic innovations. Drastic process innovations are considered because these innovations are usually granted broader protection by the Patent Office (EPO 2000a,

[1] An example is the Scripps Clinic & Research Found v. Genentech court case. In this case the court found that a blood clotting protein made with recombinant DNA techniques violated a product patent held by Scripps Clinic who had purified the same protein from human blood. Even though Genentech's process was more efficient, it could not be used because Scripps Clinic held a product patent (Merges and Nelson 1990).

USPTO 1999), as has been explained in chapters II and IV. Given that broad patents are challenged and invalidated more often than narrow ones (Waterson 1990, Lanjouw and Schankerman 2001, Merges and Nelson 1990), the innovator of a drastic process should be more careful when he determines the breadth of patent protection claimed as he cannot depend on the Patent Office to structure his claims. In addition, drastic process innovations are associated with greater innovation rents, which increase the incentive of other parties to challenge the validity of the patent and to litigate (Harhoff and Reitzig 2000, Lanjouw and Schankerman 2001).

The patent validity challenges can be either direct or indirect. Direct validity challenges are launched in the Patent Office or in the courts and they usually take place shortly after the issue of the patent. Indirect validity challenges are launched in the courts during an infringement trial.[2] This chapter considers only direct validity challenges.[3] This chapter examines the determination of the optimal patent breadth for the innovator when he is faced with a positive probability that the validity of his patent will be directly challenged. The assumption made in the patent literature (i.e., Merges and Nelson 1990, Gilbert and Shapiro 1990, Lanjouw and Schankerman 2001) that the patentee always applies for the broadest scope of patent protection is explicitly examined in this chapter.

The rest of the chapter is organized as follows. Section 5.2 presents the theoretical framework developed to examine the innovator's strategic patenting behavior. The main

---

[2] As it has been explained in chapter II, it is a common practice of accused infringers to respond to accusations of infringement with a counterclaim that the patent is invalid and it should be revoked.
[3] Indirect validity challenges are not considered because, unlike chapter IV, this chapter does not examine patent infringement.

assumptions of the model are stated in this section. Section 5.3 describes the analytical solution of the strategic patent breadth model. Finally, section 5.4 concludes the chapter.

5.2 Theoretical Development of the Strategic Patent Breadth Model

The determination of the optimal patent breadth for a strategically behaving innovator who has invented a process innovation and seeks patent protection is modeled as a sequential game of complete information. The agents involved in the game are an incumbent/patentee who is a holder of a process patent and an entrant who considers entering into the incumbent's market. It is assumed that the process innovation that is generated by the incumbent meets the patentability requirements and that the regulator (i.e., Patent Office) always grants the patent as claimed. As in chapter IV the regulator is not explicitly modeled. The model considers the determination of the strategically optimal patent breadth when the innovator has no guidance from the Patent Office in structuring his claims.

The game consists of two periods, period one, which takes place over the time interval T_G–T_0, and period two, which takes place over the time interval T_0–∞. The expression T_G denotes the time that the patent is granted and the expression T_0 (T_0=0) the time that the incumbent markets the new product and the entrant enters the market. During the first period of the game, the incumbent, having developed a process innovation and having decided to seek patent protection, determines the breadth, b, of patent protection claimed. During this period the validity of the patent may be directly challenged. The outcome of the challenge determines whether the validity of the patent is upheld or not. If the validity challenge is successful and the patent is revoked, the entrant enters the market at time T_0 using the incumbent's process and the entrant and the incumbent choose their respective output levels and compete in the market. If the patent validity is not challenged or if it is challenged and the challenge is unsuccessful (i.e., the patent is found to be valid), then the entrant, starting at time T_0, determines the flow of R&D spending, x, that will enable her to generate a non-infringing process. The incumbent operates as a monopolist for as long as the entrant is not successful in generating the non-infringing process. Once the entrant succeeds in generating the non-infringing process, however, the incumbent and the entrant choose their respective output levels and compete in the market.

The single entrant assumption is made to simplify the analysis. The assumption implies that either there is a minimum efficient scale requirement in this industry or that large sunk costs not linked to the R&D process need to be incurred upon entry that prevent the market from becoming competitive even when the incumbent's patent is revoked. Thus, the sunk costs that need to be incurred by the players upon entry are exogenous (the level of the sunk costs is not affected by the players' decisions, e.g., regulatory costs)[4] and their level is such that $\Pi_{(n=2)} \geq 0$ while $\Pi_{(n=3)} < 0$ (n denotes the number of players).

The incumbent's decision to invest in R&D and patent his product is not considered in this game. The above decisions are treated as exogenous. The only decision the incumbent makes is to determine the breadth of patent protection for his process. The length of patent protection is assumed to be fixed and for simplicity it is also assumed to be infinite. Thus, the patent will stay active unless it is invalidated

[4] In contrast, the level of endogenous sunk costs is determined by the players' decisions, e.g., advertising or R&D expenditure.

during a patent validity challenge and is thus revoked. It is also assumed that the incumbent's patent does not infringe on any previous product or process patent and there is only one Patent Office where the incumbent can apply for patent protection. Time is modeled as being continuous and complete and perfect information are assumed. The incumbent acts strategically taking into consideration the entrant's response to different patent breadth choices when he determines the breadth of patent protection claimed.

A summary of the formal strategic patent breadth determination game is depicted diagrammatically in Figure 5.1. In period one the incumbent determines the breadth of patent protection claimed, b, and he is granted a process patent. The patent is then challenged by a third party with probability δ and during the challenge process the viability of the patent is determined. The patent is upheld with probability μ and it is revoked with probability $1-\mu$. The upholding or revoking of the patent marks the end of period one. In period two the product is marketed by the incumbent. If the validity of the patent is not challenged or if it is challenged and upheld, then at the beginning of period two the entrant chooses the optimal flow of R&D spending, x. The incumbent operates as a monopolist for as long as the entrant is not successful in generating her own non-infringing process. Once the entrant succeeds, however, the incumbent and the entrant choose their respective output levels and they each earn duopoly profits. The payoffs for the incumbent and the entrant when the patent is challenged and upheld are given by $E(\Pi_I)_U^C$ and $E(\Pi_E)_U^C$, respectively (see payoffs at A). If the patent is not challenged, the payoffs are given by $E(\Pi_I)^{NC}$ and $E(\Pi_E)^{NC}$, respectively (see payoffs at C). If the patent is revoked after it has been challenged, then starting at the beginning of period two the entrant produces the new non-patentable product using the incumbent's process and the incumbent and the entrant receive payoffs $E(\Pi_I)_R^C$ and $(\Pi_E)_R^C$, respectively (see payoffs at B).

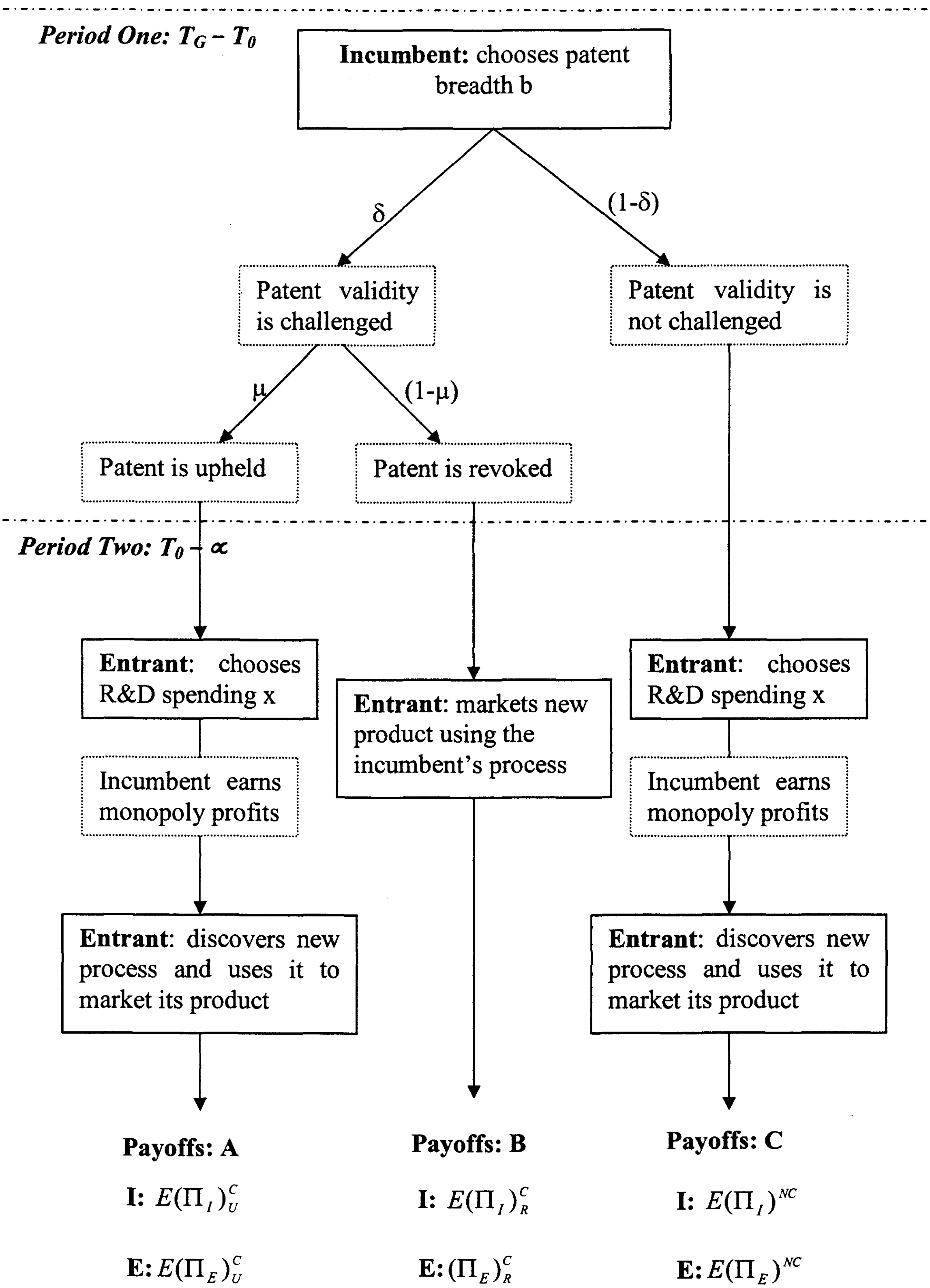


Figure 5.1 The Strategic Patent Breadth Game

In this model the entrant knows (with certainty) the outcome of the patent challenge when she decides on the level of R&D expenses to be incurred or when she decides on whether to enter the market using the incumbent's process. The incumbent on the other hand knows only the probability with which his patent will be challenged and the probability with which his patent will be upheld when challenged when he determines the breadth of patent protection claimed.

5.2.1 The Process Innovation Space, Patent Breadth and the R&D Process

It is assumed that the patentable process is used for the production of a new non-patentable product (e.g., the t-PA drug in U.K.). The potential entrant is thus free to produce the new product by generating her own non-infringing process. This model assumes that if the entrant enters the market she will do so without infringing the incumbent's process. In addition, if the entrant is successful in generating her own process, she does not have to patent it since further entry is not anticipated.

The model assumes that the patented process results in zero per unit production costs and that if the entrant succeeds in generating her own non-infringing process, her process will be equally efficient (i.e., the entrant's process also results in per unit production costs of zero). Thus, the model is not a quality ladder model where one innovator supersedes the other in producing a better innovation. Both the incumbent/patentee and the potential entrant use their processes to produce the new non-patentable product, which is viewed by the consumers as a homogeneous product. In other words, consumers are indifferent as to whether the new product was made with the incumbent's or the entrant's process.

It is also assumed that there are no close substitutes for the new product. The above assumption implies that the incumbent will make monopoly profits for as long as the entrant fails to generate a non-infringing process. Once the entrant succeeds in generating her own non-infringing process, the incumbent and the entrant will share the market, each making duopoly profits.

The process innovation space and the breadth of patent protection are depicted in Figure 5.2. The line of unit length represents the process innovation space. Each point within this space represents a process that is capable of producing the product in question at the same production cost. Thus, different points on the unit length line refer to the different processes that can be used to produce the non-patentable product at a per unit cost of zero. The closer are two points in the process innovation space, the more similar are the processes in terms of the way they work in generating the given product.

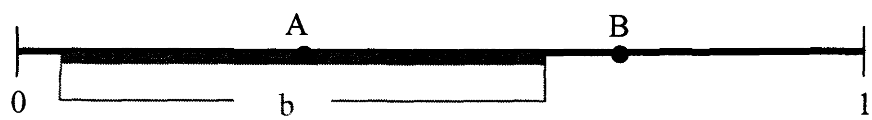


Figure 5.2 The Process Innovation Space and the Breadth of Patent Protection

Point A in Figure 5.2 refers to the patented process generated by the incumbent. Patent breadth refers to the area on the unit length line around point A which is protected by the patent. Patent breadth includes all the processes that, if they were developed by competitors, would infringe the patent. Patent breadth takes values in the interval $b \in (0,1]$. A patent breadth value close to zero ($b \to 0$) implies that the patent protects only against duplication of the patented process. On the other hand, a patent breadth value equal to one (b=1) implies that there is no other process that can be used to produce the non-patentable product without infringing the patent. As in chapter IV, it is assumed that patent breadth defines an exact border of protection (i.e., fencepost patent system).

To simplify the model it is assumed that it is a third party and not the potential entrant who directly challenges the validity of the patent.[5] Thus, in this model, the entrant benefits from the validity challenge without incurring the opposition costs. The costs incurred by the incumbent during a validity challenge are denoted by C_T and are assumed to be independent of the breadth of patent protection. In addition, it is assumed that the incumbent's opposition costs do not affect the probability that the patent will be challenged and the probability that the validity of the patent will be upheld.

The probability that the validity of the patent will be directly challenged is denoted by δ and it is a function of patent breadth. Recent empirical studies have found a positive relationship between the breadth of the patent, measured by the number of claims made, and the probability of validity challenges (Lanjouw and Schankerman 2001). In addition, Lentz (1988) and Merges and Nelson (1990) observe that the greater is the breadth of patent protection, the greater is the probability that the validity of the patent will be challenged. Following the above studies, this model assumes a positive relationship between patent breadth and the probability that the validity of the patent will be challenged. For simplicity, it is further assumed that when the maximum patent breadth is claimed (b_{max}=1), the validity of the patent is always challenged. These

[5] Third parties are allowed to challenge the validity of patents in the Patent Office without having to prove any special interest for doing so (see chapter II). Harhoff and Reitzig (2000) state that various interest groups are trying to influence the European patenting practice by filing opposition cases especially against biotechnology patents.

assumptions are captured by assuming that the probability that the validity of the patent will be directly challenged, δ, is equal to $\delta = b$.

The patent may not always be found valid during the patent validity challenge. There is a probability, denoted by μ, that the validity of the patent will be upheld during the validity challenge, where μ is given by $\mu = 1 - b$. Thus, the greater is the breadth of patent protection, the smaller is the probability that the validity of the patent will be upheld. As it has been discussed in chapter IV, the above assumption is justified by the fact that the greater is patent breadth, the harder it is to show novelty, nonobviousness and enablement. In addition, empirical evidence suggests that courts tend to uphold narrow patents and revoke broad ones (Merges and Nelson 1990, Waterson 1990). As in chapter IV, when patent breadth takes its maximum value ($b_{max}=1$), the patent is always found to be invalid (μ=0).

In statistical terms, the event that the patent will be challenged and the event that the validity of the patent will be upheld are treated as independent.[6] This assumption is valid given that the patent validity challenger is not the one who rules on whether the patent is valid. There is no evidence in the literature to suggest that there is a relationship between the probability that the patent will be challenged and the way the courts and/or the Patent Office rule on patent validity issues.

When the validity of the patent is not challenged, or when it is challenged and upheld, the entrant must invest in R&D to generate her own non-infringing process to produce the non-patentable product, if she wants to enter the market. To capture the

---

[6] This assumption implies that the probability that the patent will be upheld given that is has been challenged is equal to the probability that the patent will be upheld, i.e., prob[μ | δ]=prob[μ].

uncertainty associated with the R&D process it is assumed that the innovation process is stochastic. Innovation in this model occurs according to the Poisson process.[7] The research technology is 'memoryless'; that is, the probability that the entrant will succeed in generating an innovation at any given point in time depends only on the current R&D expenditure, not on past R&D experience (Tirole 1988). This is a common assumption in the R&D literature and is made to simplify the analysis (Loury 1979, Lee and Wild 1980). The instantaneous probability of success is denoted by λ and is constant. The parameter λ shows that if the entrant has not succeeded by time τ in generating a non-infringing process then the probability of succeeding at the next instant, that is at $\tau + dt$, is λdt. The elapsed time, τ, before an innovation arrives has a probability density function described by the exponential distribution $f(\tau) = \lambda e^{-\lambda\tau}$ for $\lambda>0$ and $0\leq\tau\leq\infty$ and a cumulative probability function $F(\tau) = 1 - e^{-\lambda\tau}$. The cumulative distribution gives the probability that success will occur by time τ (i.e., $F(\tau) = prob[t \leq \tau]$).

In this model it is assumed that the instantaneous probability of success λ is a function of the entrant's R&D spending per unit of time, denoted by x, and the breadth of patent protection b, $\lambda = f(x,b)$. The flow rate of R&D spending, x, is assumed to be constant and it is incurred by the entrant for as long as it takes to realize a success. Following standard economic theory assumptions, it is assumed that the R&D spending

[7] The Poisson distribution gives the probability of x occurrences of an event over a specified period of time. The probability mass function is given by $P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$ for x=0,1,2….., and $\lambda>0$ and by $P(x) = 0$ otherwise. The parameter λ gives the mean rate of occurrences per unit of time and it also represents the variance of the Poisson distribution.

per unit of time increases the probability of success at a decreasing rate; $\lambda_x>0$, $\lambda_{xx}<0$ and also $\ell im_{x\to\infty} \lambda_x = 0$ and $\lambda_x(0)\to\infty$ (Loury 1979, Reinganum 1983).

The instantaneous probability of success λ is also a function of the breadth of patent protection since in this model success implies that the entrant will only be able to enter with a non-infringing process. That is, the entrant must generate a process outside the technological territory – i.e., the patent breadth – claimed by the incumbent. Given that the entrant has not already succeeded, it is assumed that the greater is the patent breadth, the smaller is the probability that the entrant will succeed at the next instant, in generating a non-infringing process for producing the new non-patentable product. It is thus assumed that the breadth of patent protection decreases the probability of success at an increasing rate; $\lambda_b<0$, $\lambda_{bb}>0$. The justification for this assumption is that since the entrant will enter with a non-infringing process, the greater is patent breadth the more dissimilar will be the two processes – the further away from the patentee's process the entrant's process will be in the process innovation space in Figure 5.2. This, in turn, implies that the greater is patent breadth, the less useful is the information disclosed by the patent for the entrant and thus the more difficult it becomes for the entrant to generate her non-infringing process.

To completely describe the instantaneous probability of success, λ, the instantaneous probability of success is assumed to be either additively or multiplicatively separable in the flow of R&D spending and in patent breadth, i.e., $\lambda = \varphi(x) + \psi(b)$ or $\lambda = \varphi(x) \cdot \psi(b)$. The functions $\varphi(x)$ and $\psi(b)$ satisfy all theoretical assumptions concerning the instantaneous probability of success, that is, $\varphi_x > 0$,

$\varphi_{xx} < 0$, $\psi_b < 0$ and $\psi_{bb} > 0$. With the additively separable formulation, the marginal effect of R&D spending on the probability of success is independent of the breadth of the patent, $\lambda_{xb} = 0$. With the multiplicatively separable formulation the marginal effect of R&D spending on the probability of success is inversely related to the breadth of the patent, $\lambda_{xb} < 0$ (see Proposition 5.2 for a formal proof).

Given the above, when $\lambda = \varphi(x) + \psi(b)$ the incumbent's patent breadth choice affects the entrant's probability of success, λ, only directly ($\lambda_b < 0$). When $\lambda = \varphi(x) \cdot \psi(b)$ the incumbent's patent breadth choice affects the entrant's probability of success, λ, both directly ($\lambda_b < 0$) and indirectly ($\lambda_{xb} < 0$). In this case, as patent breadth increases, the harder it becomes to generate a non-infringing process (i.e., direct effect) and the less effective R&D spending becomes in increasing the probability of success (i.e., indirect effect). An additively separable function and a multiplicatively separable function that satisfy all theoretical assumptions regarding the instantaneous probability of success are given by $f_1 : \lambda = x^\theta + \frac{1}{b}$ and $f_2 : \lambda = \frac{x^\theta}{b}$ respectively, where $\theta \in (0,1]$.

5.3 The Analytical Solution of the Strategic Patent Breadth Game

Given the assumption of complete and perfect information, the incumbent knows when he determines the breadth of patent protection claimed how patent breadth affects the probability that the patent will be challenged, δ, the probability that the validity of the patent will be upheld after challenge, μ, and the entrant's probability of succeeding at any given instant in generating a non-infringing process, λ. The incumbent chooses the breadth of patent protection that will induce the desired behavior from the entrant and will allow him to maximize the rents that he can appropriate from his innovation.

The optimal breadth of patent protection for the innovator is determined using backwards induction. Backwards induction yields the only subgame perfect Nash equilibrium of this game by eliminating Nash equilibria that do not represent credible threats (Fudenberg and Tirole 1991). The duopoly profits that are realized at the second period of the game when both the incumbent and the entrant operate in the market are determined first. The entrant's decision of the optimal R&D spending is determined next and the incumbent's optimal patent breadth choice is determined last.

5.3.1 Determination of the Duopoly Profits

During the second period of the game (T_0–∞) both the incumbent and the entrant produce the product when either the entrant succeeds in generating the non-infringing process or when the patent is revoked after being challenged. Since production costs have been assumed to be zero, both players will produce the same output and will earn the same rate of instantaneous profits. These instant profits are given by $\Pi_I = \Pi_E = \Pi_d > 0$.[8] Although the entrant earns the same level of instantaneous profits the discounted profits will differ from those of the incumbent depending on the R&D expenditures and on the exogenous sunk costs she has to incur.

---

[8] Given that the incumbent and the entrant compete in quantities and not in prices and that the unit production costs are zero the duopoly profits that they realize are positive. This is a standard result in the literature when firms with the same cost structure engage in Cournot competition (Tirole 1988, Carlton and Perloff 1999).

5.3.2 The Entrant's Optimal R&D Spending Decision

Two cases emerge regarding the entrant's behavior depending on whether the patent is challenged and revoked or on whether the patent is not challenged or is challenged and upheld. The entrant's optimal decision when the patent is challenged and revoked is examined first.

- *The Patent is Challenged and Revoked*

The entrant does not have to make an investment decision if the patent is revoked after being challenged. Since generating a new process is costly for the entrant (i.e., positive R&D costs are required), the entrant simply uses the incumbent's process to produce the new product. When the entrant uses the incumbent's process to produce the new product her discounted profits are given by:

$$(\Pi_E)^R = \int_0^\infty e^{-rt}\Pi_d dt - F = \frac{\Pi_d}{r} - F \tag{5.1}$$

where r is the discount rate and F are the exogenous sunk costs incurred by the entrant at time zero (T_0).

The entrant finds it optimal to enter the market when the patent is challenged and revoked if $(\Pi_E)^R > 0 \Rightarrow \frac{\Pi_d}{r} > F$.

- *The Patent is Not Challenged or is Challenged and Upheld*

When the patent is not challenged, or is challenged and upheld, the entrant must decide on the flow of R&D spending that will enable her to generate the non-infringing process that will be used for the production of the new product. The entrant chooses the flow of R&D spending, x, that maximizes the present value of her expected profits. Note that the entrant's expected profits are the same irrespective of whether the patent is not challenged or challenged and upheld ($E(\Pi_E)_U^C = E(\Pi_E)^{NC}$) since it is not the entrant but a third party that challenges the validity of the patent. The entrant's objective function is given by:

$$\max_x E(\Pi_E)_U^C = E(\Pi_E)^{NC} = \int_0^\infty e^{-rt} e^{-\lambda(x,b)t}(\lambda(x,b)\Pi_d - x)dt - F \tag{5.2}$$

Equation (5.2) shows that if the entrant has not succeeded before time t in generating the non-infringing process, she then receives Π_d if she succeeds at time t. This event has probability density $\lambda(x,b)e^{-\lambda(x,b)t}$. The entrant pays R&D costs of x so long as no success has occurred. This event has probability $e^{-\lambda(x,b)t}$. Finally, the entrant pays costs F at time zero irrespective of whether she succeeds in generating the process.

Performing the indicated integrations the entrant's objective function can be expressed as:

$$\max_x E(\Pi_E)_U^C = E(\Pi_E)^{NC} = \frac{\lambda(x,b)\Pi_d - x}{r + \lambda(x,b)} - F \tag{5.3}$$

The entrant chooses the flow of R&D spending that maximizes her objective function given in equation (5.3). Optimization of equation (5.3) yields the following first order conditions (F.O.C.) for a maximum:

$$\frac{\partial E(\Pi_E)_U^C}{\partial x} = \frac{\partial E(\Pi_E)^{NC}}{\partial x} = 0 \Rightarrow x - \frac{r + \lambda(x,b)}{\lambda_x} + r\Pi_d = 0 \Rightarrow x^* = x(b,r,\Pi_d) \tag{5.4}$$

The F.O.C. yield the optimal flow of R&D spending expressed in terms of known parameters; the breadth of patent protection, the duopoly profits and the discount rate. The F.O.C. implicitly define the entrant's best response function, which shows how the entrant responds to different patent breadth choices.

To graphically characterize the optimal level of the flow of R&D spending let $g(x) = x$ and $h(x) = \frac{r + \lambda(x,b)}{\lambda_x} - r\Pi_d$. The F.O.C. can then be written as:

$$g(x) - h(x) = 0 \Rightarrow x = \frac{r + \lambda(x,b)}{\lambda_x} - r\Pi_d \qquad (5.5)$$

The second order conditions (S.O.C.) imply that for a maximum the condition given in equation (5.6) must be satisfied.

$$g_x - h_x < 0 \Rightarrow g_x < h_x \qquad (5.6)$$

Equation (5.6) shows that at the optimum the slope of $h(x)$ must be greater than the slope of $g(x)$; $h(x)$ must cut $g(x)$ from below at the optimum. Given that $g_x = 1 > 0$ equation (5.6) implies that $h(x)$ must be increasing in x and also that $h_x > 1$. It is easily verified that both the above conditions for the existence of an optimum are satisfied as $h_x = \frac{\lambda_x \lambda_x}{(\lambda_x)^2} - \frac{\lambda_{xx}(r+\lambda)}{(\lambda_x)^2} = 1 - \frac{\lambda_{xx}(r+\lambda)}{(\lambda_x)^2} > 1$, since $\lambda_{xx} < 0$. The S.O.C. are satisfied for both the additive and the multiplicative formulations of the instantaneous probability of success, λ. Also, $h(0) = -r\Pi_d$ since $\lambda_x(0) \to \infty$ which holds due to the theoretical properties of the instantaneous probability of success.

The slope of $h(x)$ is decreasing in the flow of R&D spending, x, for both the additive and the multiplicative formulations of the instantaneous probability of success, λ, that is, $h_{xx} = -\frac{\lambda_{xxx}(r+\lambda) + \lambda_x \lambda_{xx}}{(\lambda_x)^2} + \frac{2(\lambda_{xx})^2(r+\lambda)}{(\lambda_x)^3} \le 0$. A formal proof is presented in the Appendix. Note that the determination of the curvature of $h(x)$ is not important for the results, it is necessary only for the graphical representation of the optimum.

Figure 5.3 depicts the graphical representation of the determination of the optimal flow of R&D spending. In Figure 5.3 the curve $h(x)$ cuts the curve $g(x)$ at two different points. From these two points only point A is an optimum because only at point A is the slope of $h(x)$ greater than the slope of $g(x)$.

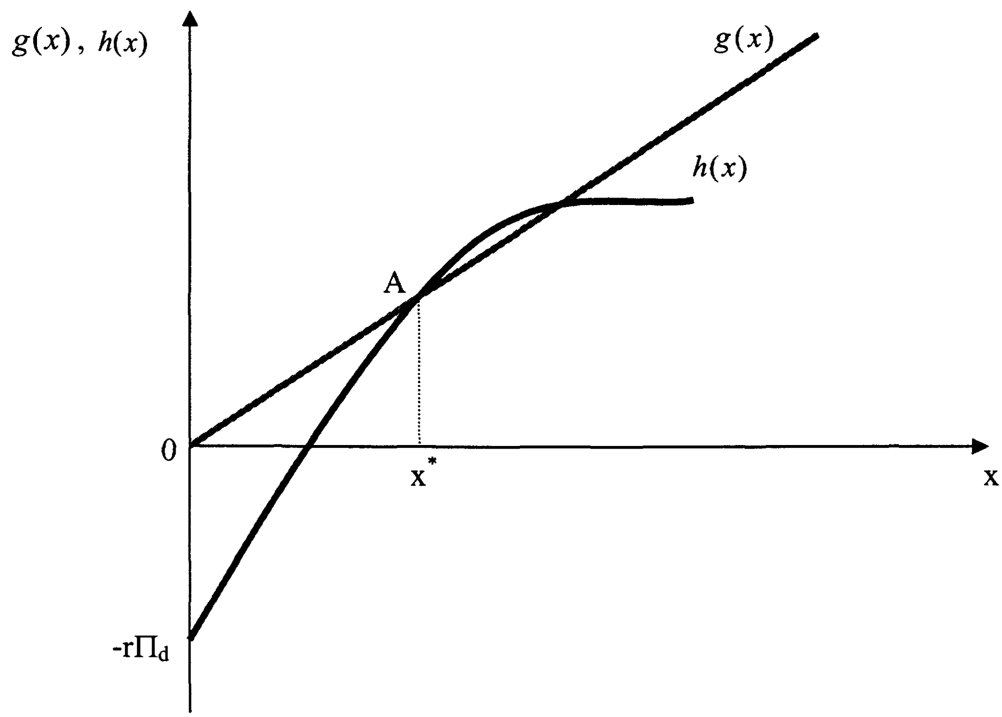


Figure 5.3 Graphical Representation of the Determination of the Optimal Flow of R&D Spending (x^*).

The entrant's expected profits when the patent is not challenged or when it is challenged and upheld are obtained by substituting the optimal flow of R&D spending into the entrant's expected profit function. The substitution yields the payoffs given by equation (5.7).

$$E(\Pi_E)_U^C = E(\Pi_E)^{NC} = \frac{\lambda(x^*,b)\Pi_d - x^*}{r+\lambda(x^*,b)} - F \tag{5.7}$$

The entrant will enter the market only if she realizes positive profits, that is, if $E(\Pi_E)_U^C = E(\Pi_E)^{NC} > 0$. This condition can be expressed as:

$$\frac{\lambda(x^*,b)\Pi_d - x^*}{r+\lambda(x^*,b)} - F > 0 \Rightarrow \Pi_d > \frac{x^* + (r+\lambda(x^*,b))F}{\lambda(x^*,b)} \tag{5.8}$$

Note that the entry condition in equation (5.8) is determined by the level of duopoly profits, the discount rate, the exogenous sunk costs and the incumbent's patent breadth decision. Equation (5.8) opens the possibility that a patent breadth value $\hat{b} \in (0,1]$ may exist such that the entry condition is not satisfied. If $\hat{b}$ exists and it is chosen by the incumbent, then the entrant will not enter the market.

When entry is not deterred the entrant's optimal flow of R&D spending is given by equation (5.4). This equation can be used to determine the effect of a change in the breadth of patent protection, the level of duopoly profits and the discount rate, on the optimal flow of R&D spending.

The effect of a change in patent breadth on the optimal flow of R&D spending is determined by examining $\frac{dx^*}{db}$. The expression for the term $\frac{dx^*}{db}$ is derived by totally differentiating the optimality condition $g(x) = h(x) \Rightarrow x = \frac{r+\lambda}{\lambda_x} - r\Pi_d$ (i.e., equation (5.5)), with respect to the optimal flow of R&D spending, x^*, and patent breadth, b. The result of this differentiation is:

$$\frac{\partial g(x^*)}{\partial x^*}dx^* + \frac{\partial g(x^*)}{\partial b}db = \frac{\partial h(x^*)}{\partial x^*}dx^* + \frac{\partial h(x^*)}{\partial b}db \Rightarrow$$
$$(g_x - h_x)dx^* = (h_b - g_b)db \Rightarrow \frac{dx^*}{db} = \frac{h_b - g_b}{(g_x - h_x)} \qquad (5.9)$$

It is known from equation (5.6) that $g_x - h_x < 0$. Also, $g_b = 0$. Thus, the sign of the slope of the best response curve, $\frac{dx^*}{db}$, depends on the sign of the term h_b, where $h_b = \frac{\lambda_b \lambda_x - \lambda_{xb}(r+\lambda)}{(\lambda_x)^2}$.

The nature of the instantaneous probability of success, λ, must be specified before the relationship between the optimal level of R&D spending, x^*, and patent breadth, b, can be determined. This is so because the term h_b depends on the term λ_{xb}, the sign of which depends on whether the instantaneous probability of success, λ, is additively or multiplicatively separable in the flow of R&D spending, x, and in patent breadth, b.

Proposition 5.1 *When the instantaneous probability of success, λ, is additively separable in patent breadth, b, and in the flow of R&D spending, x, (i.e., $\lambda = \varphi(x) + \psi(b)$), the effect of a change in patent breadth on the optimal flow of R&D spending is positive, i.e.,* $\frac{dx^*}{db} > 0$.

Proof:

When $\lambda = \varphi(x) + \psi(b)$, then $\frac{\partial \lambda_x}{\partial b} = \lambda_{xb} = 0$. Given that $\lambda_b < 0$ and $\lambda_x > 0$, $h_b = \frac{\lambda_b \lambda_x - \lambda_{xb}(r+\lambda)}{(\lambda_x)^2}$<0. The slope of the best response function, given by equation (5.9), is thus positive, i.e., $\frac{dx^*}{db} > 0$. □

The above result suggests that as patent breadth increases so does the flow of R&D spending. The intuition behind this result is as follows. The entrant responds to an increase in patent breadth with an increase in her flow of R&D spending trying to counterbalance the negative effect that the increase in patent breadth has on the probability of success.

Proposition 5.2 *When the instantaneous probability of success, λ, is multiplicatively separable in the patent breadth, b, and in the flow of R&D spending, x, (i.e., $\lambda = \varphi(x) \cdot \psi(b)$) the effect of a change in patent breadth on the optimal flow of R&D spending is negative, i.e.,* $\frac{dx^*}{db} < 0$.

Proof:

When $\lambda = \varphi(x) \cdot \psi(b)$, then $\lambda_{xb} < 0$, since $\frac{\partial \lambda}{\partial x} = \varphi_x \psi(b)$, $\frac{\partial \lambda}{\partial b} = \varphi(x)\psi_b$ and $\frac{\partial \lambda_x}{\partial b} = \frac{\partial \lambda_b}{\partial x} = \lambda_{xb} = \varphi_x \psi_b < 0$. Thus, patent breadth affects the probability of success both directly ($\lambda_b < 0$) and indirectly ($\lambda_{xb} < 0$). The sign of the term h_b is positive,

$h_b = \frac{\lambda_b \lambda_x - \lambda_{xb}(r+\lambda)}{(\lambda_x)^2} = -\frac{r}{\varphi_x \psi_b} > 0$. Given that $\frac{dx}{db} = \frac{h_b}{(g_x - h_x)}$ (from equation 5.9) and $g_x - h_x < 0$, the slope of the best response function is negative, i.e., $\frac{dx^*}{db} < 0$. □

The intuition behind the result presented in Proposition 5.2 is as follows. An increase in patent breadth has two effects on the entrant. First, she knows that a change in b means she will have to spend more to be able to succeed (to counterbalance the negative effect that an increase in the patent breadth has on the probability of success). Second, she also knows that the effect of the additional R&D spending on the probability of success will now be smaller (due to $\lambda_{xb} < 0$). Since an increase in patent breadth makes investment less efficient and more costly for the entrant, the entrant responds with a reduction in the optimal flow of R&D spending.

Having determined how patent breadth affects the flow of R&D spending under different assumptions about the relationship between the flow of R&D spending and patent breadth (i.e., $\lambda_{xb} = 0$ and $\lambda_{xb} < 0$), the effect of a change in patent breadth on the total expected R&D costs that the entrant must incur before a success is realized can be determined. The total expected R&D costs to be incurred by the entrant are given by:

$$TEC_E = \tau_e x = \frac{1}{\lambda(x,b)} x \tag{5.10}$$

In equation (5.10), $\tau_e = \frac{1}{\lambda(x,b)}$ denotes the average elapsed time before success is realized; this is the mean of the exponential distribution $f(\tau) = \lambda e^{-\lambda\tau}$. The average elapsed time before success occurs is decreasing in the flow of the R&D spending,

$\frac{\partial \tau_e}{\partial x} = \frac{\partial \tau_e}{\partial \lambda}\frac{\partial \lambda}{\partial x} = -\frac{1}{\lambda^2}\lambda_x < 0$ and increasing in the breadth of patent protection, $\frac{\partial \tau_e}{\partial b} = \frac{\partial \tau_e}{\partial \lambda}\frac{\partial \lambda}{\partial b} = -\frac{1}{\lambda^2}\lambda_b > 0$. Thus, on the one hand, the greater is the flow of R&D spending, the greater is the probability that success will be realized the next instant, and the shorter is the time that elapses before success occurs. On the other hand, the greater is patent breadth, the smaller is the probability that success will occur the next instant and thus the longer is the period that elapses before success occurs. The propositions that follow describe the relationship between patent breadth and the total expected R&D costs when the instantaneous probability of success is additively and multiplicatively separable in the flow of R&D spending and in patent breadth.

Proposition 5.3 *The total expected R&D costs are increasing in patent breadth,* $\frac{dTEC_E}{db} > 0$, *when the instantaneous probability of success,* λ, *is additively separable in patent breadth, b, and in the flow of R&D spending, x.*

Proof:

When $\lambda = \varphi(x) + \psi(b)$, $\frac{\partial TEC_E}{\partial b} = \frac{\partial \tau_e}{\partial b}x + \tau_e x_b = \frac{x_b\lambda - \lambda_b x}{\lambda^2} > 0$ since $x_b > 0$ as shown in Proposition 5.1 and $\lambda_b < 0$. □

The intuition behind the result presented in Proposition 5.3 is as follows. When the instantaneous probability of success is additively separable in the flow of R&D spending and in patent breadth, patent breadth affects the expected total R&D costs in two ways. First, a higher patent breadth increases the elapsed time before success is realized ($\frac{\partial \tau_e}{\partial b} > 0$), and second, greater patent breadth increases the flow of R&D spending ($x_b > 0$). Both outcomes imply that the expected total R&D costs to be incurred by the entrant are rising in b.

Proposition 5.4 *The effect of an increase in patent breadth on the total expected R&D costs when the instantaneous probability of success, λ, is multiplicatively separable in patent breadth, b, and in the flow of R&D spending, x, depends on whether an increase in patent breadth increases the elapsed time before success is realized more than it decreases the flow of R&D spending.*

Proof:

The effect of a change in patent breadth on the total expected R&D costs is given by $\frac{\partial TEC_E}{\partial b} = \frac{\partial \tau_e}{\partial b} x + \tau_e x_b = \frac{x_b \lambda - \lambda_b x}{\lambda^2}$. When $\lambda = \varphi(x) \cdot \psi(b)$ then $x_b < 0$ as shown in Proposition 5.2 and $\lambda_b < 0$. Given the above, the sign of $\frac{dTEC_E}{db}$ depends on the relative magnitudes of the terms $\frac{\partial \tau_e}{\partial b} x$ and $\tau_e x_b$, which are positive and negative respectively (or equivalently the terms $x_b \lambda$ and $\lambda_b x$, which are both negative). When $\tau_e x_b > \frac{\partial \tau_e}{\partial b} x$ then $\frac{dTEC_E}{db} < 0$ while when $\tau_e x_b < \frac{\partial \tau_e}{\partial b} x$ then $\frac{dTEC_E}{db} > 0$. □

When the instantaneous probability of success is multiplicatively separable in the flow of R&D spending and in patent breadth, patent breadth affects the expected total R&D costs in two countervailing ways. On the one hand, an increase in patent breadth increases the elapsed time before success is realized; on the other hand, an increase in patent breadth decreases the flow of R&D spending. When $\frac{dTEC_E}{db} > 0$ then even though the amount spend on R&D per unit of time decreases as patent breadth increases this amount is now spread over a longer period making the total effect of the increase in patent breadth positive. When $\frac{dTEC_E}{db} < 0$ then even though the period over which the flow R&D costs are incurred increases as patent breadth increases, the decrease in the flow of R&D spending (caused by the patent breadth increase) is greater in absolute terms making the total effect of an increase in patent breadth on the total expected R&D costs negative.

The effect of a change in the anticipated level of duopoly profits on the flow of the R&D spending is determined by totally differentiating the optimality condition, $g(x) = h(x) \Rightarrow x = \frac{r+\lambda}{\lambda_x} - r\Pi_d$, with respect to the optimal flow of R&D spending, x* and the anticipated duopoly profits, Π_d. The expression for $\frac{dx^*}{d\Pi_d}$ is given by:

$$
\begin{aligned}
&\frac{\partial g(x^*)}{\partial x^*}dx^* + \frac{\partial g(x^*)}{\partial \Pi_d}d\Pi_d = \frac{\partial h(x^*)}{\partial x^*}dx + \frac{\partial h(x^*)}{\partial \Pi_d}d\Pi_d \Rightarrow \\
&(g_x - h_x)dx^* = (h_{\Pi_d} - g_{\Pi_d})d\Pi_d \Rightarrow \frac{dx^*}{d\Pi_d} = \frac{h_{\Pi_d} - g_{\Pi_d}}{(g_x - h_x)}
\end{aligned}
\tag{5.11}
$$

Proposition 5.5 *The optimal level of the flow of R&D spending is increasing in the duopoly profits that the entrant anticipates to make if she succeeds in generating a non-*

infringing process, $\frac{dx^*}{d\Pi_d} > 0$, *for both the additive and the multiplicative formulations of the instantaneous probability of success,* λ.

Proof:

From the S.O.C., $g_x - h_x < 0$. Also, $g_{\Pi_d} = 0$ and $h_{\Pi_d} = -r < 0$, which implies that $\frac{dx^*}{d\Pi_d} > 0$. Thus, as it would be expected, the entrant reacts to an increase in the anticipated duopoly profits by increasing the optimal flow of R&D spending. □

Proposition 5.6 *The total expected R&D costs that are incurred by the entrant before a success occurs are increasing in the duopoly profits that the entrant anticipates to make if she succeeds in generating a non-infringing process,* $\frac{dTEC_E}{d\Pi_d} > 0$.

Proof:

The change in the total expected R&D costs that follows a change in the anticipated duopoly profits is given by $\frac{\partial E(TC_E)}{\partial \Pi_d} = (\frac{\partial \tau_e}{\partial \lambda}\frac{\partial \lambda}{\partial x}\frac{\partial x}{\partial \Pi_d})x + \frac{\partial x}{\partial \Pi_d}\tau_e = (\frac{-r}{g_x - h_x})(\frac{\lambda - \lambda_x x}{\lambda^2})$.

In this expression the term $(\frac{-r}{g_x - h_x}) = \frac{dx^*}{d\Pi_d}$ is positive as shown in Proposition 5.5.

The term $\frac{\lambda - \lambda_x x}{\lambda^2}$ is also positive since $\frac{\lambda}{x} > \lambda_x$. To prove the last inequality the additively and multiplicatively separable functions $f_1 : \lambda = x^\theta + \frac{1}{b}$ and $f_2 : \lambda = \frac{x^\theta}{b}$, respectively, are used. When λ is described by f_1 then

$\frac{\lambda}{x} > \lambda_x \Rightarrow x^{\theta-1} + \frac{1}{bx} > \theta x^{\theta-1} \Rightarrow 1 + \frac{1}{bx} > \theta$ which holds ∀b∈(0,1], x≥0 since θ∈(0,1).

When λ is described by f_2 then $\frac{\lambda}{x} > \lambda_x \Rightarrow \frac{x^{\theta-1}}{b} > \frac{\theta x^{\theta-1}}{b} \Rightarrow 1 > \theta$ which holds true ∀b∈(0,1], x≥0 since θ∈(0,1). □

The intuition behind the result presented in Proposition 5.6 is as follows. The increase in the anticipated duopoly profits causes the entrant to increase her flow of spending in R&D (see Proposition 5.5). The increase in the flow of R&D spending affects the expected total R&D costs in two countervailing ways. A greater flow of R&D spending directly obviously increases the expected total costs. It also indirectly decreases the expected total R&D costs by decreasing the average elapsed time before success is realized ($\frac{\partial \tau_e}{\partial x} < 0$). However, the positive direct effect is stronger than the negative indirect effect. The result is that an increase in the anticipated duopoly profits on the expected total R&D costs is positive.

The effect of a change in the discount rate on the optimal level of R&D spending is derived by totally differentiating the optimality condition $g(x) = h(x) \Rightarrow x = \frac{r+\lambda}{\lambda_x} - r\Pi_d$, with respect to the optimal flow of R&D spending, x*, and the discount rate, r. The expression for $\frac{dx^*}{dr}$ is given in equation (5.12).

$$\begin{aligned} &\frac{\partial g(x^*)}{\partial x^*} dx^* + \frac{\partial g(x^*)}{\partial r} dr = \frac{\partial h(x^*)}{\partial x^*} dx^* + \frac{\partial h(x^*)}{\partial r} dr \Rightarrow \\ &(g_x - h_x) dx^* = (h_r - g_r) dr \Rightarrow \frac{dx^*}{dr} = \frac{h_r - g_r}{(g_x - h_x)} \end{aligned} \tag{5.12}$$

From the S.O.C. it is known that $g_x - h_x < 0$. Also, $g_r = 0$. The sign of the term $h_r = \frac{1}{\lambda_x} - \Pi_d$ cannot be determined, however, without knowledge of the magnitude of the parameters in the expression. For instance, note that the impact of the discount rate on the optimal level of R&D spending depends on the magnitude of the anticipated duopoly profits, Π_d. When $\frac{1}{\lambda_x} > \Pi_d$ then an increase in the discount rate decreases the optimal level of the flow of R&D spending $\frac{dx}{dr} < 0$ while when $\frac{1}{\lambda_x} < \Pi_d$ an increase in the discount rate results in an increase in the optimal level of the flow of R&D spending $\frac{dx}{dr} > 0$. Finally, when $\frac{1}{\lambda_x} = \Pi_d$ a change in the discount rate causes no change in the optimal level of R&D spending. Given that the effect of a change in the discount rate on the optimal level of the flow of R&D spending is inconclusive, the effect of a change in the discount rate on the expected total R&D costs is also inconclusive.

5.3.3 The Incumbent's Optimal Patent Breadth Decision

Given the assumption of complete information, the incumbent knows how patent breadth affects the entrant's optimal R&D spending decision. The incumbent can then choose the breadth of patent protection that induces the desired behavior from the entrant. This is the breadth of patent protection that maximizes the incumbent's discounted expected returns.

The incumbent's expected returns are a function of his expected returns when the patent is not challenged, $E(\Pi_I)^{NC}$, and the expected returns when the patent is challenged, $E(\Pi_I)^C$. Since the incumbent's patent is challenged with probability δ and it is not challenged with probability 1-δ, the incumbent's discounted expected profits are given by equation (5.13).

$$E(\Pi_I) = \delta E(\Pi_I)^C + (1-\delta)E(\Pi_I)^{NC} \tag{5.13}$$

The incumbent's expected returns when the patent is challenged are a function of the incumbent's expected returns when the patent is challenged and upheld, $E(\Pi_I)_U^C$, and the expected returns when the patent is challenged and revoked, $E(\Pi_I)_R^C$. Given that the patent is challenged and upheld with probability μ and it is challenged and revoked with probability 1-μ, the incumbent's expected returns when the patent is challenged are given by equation (5.14). In equation (5.14) C_T denotes the legal costs incurred by the incumbent during the patent challenge process.

$$E(\Pi_I)^C = \mu E(\Pi_I)_U^C + (1-\mu)E(\Pi_I)_R^C - C_T \tag{5.14}$$

The incumbent's expected returns when his patent is not challenged or when it is challenged and upheld are the same ($E(\Pi_I)^{NC} = E(\Pi_I)_U^C$) because in both cases the incumbent operates as a monopolist until the entrant succeeds in generating a non-infringing process. Once the entrant succeeds, the incumbent shares the market with the entrant, each making duopoly profits. The incumbent's discounted expected profits when his patent is not challenged or when it is challenged and upheld are given by equation (5.15).

$$E(\Pi_I)^{NC} = E(\Pi_I)_U^C = \int_0^\infty e^{-rt} e^{-\lambda(x^*,b)t}(\Pi_m + \lambda(x^*,b)\Pi_d)dt = \frac{\Pi_m + \lambda\Pi_d}{r+\lambda} = \Pi^u \tag{5.15}$$

Equation (5.15) shows that the incumbent receives monopoly profits Π_m at t if by time t the entrant has not yet succeeded in generating a non-infringing process. This event has probability $e^{-\lambda(x^*,b)t}$. The incumbent receives duopoly profits Π_d at time t if, at t, the entrant succeeds in generating a non-infringing process. This event has a probability density function $\lambda(x^*,b)e^{-\lambda(x^*,b)t}$.

When the patent is challenged and revoked, the entrant enters using the incumbent's process and the incumbent shares the market with the entrant making duopoly profits. The incumbent's discounted profits when his patent is challenged and revoked are given by equation (5.16).

$$(\Pi_I)_R^C = \int_0^\infty e^{-rt}\Pi_d dt = \frac{\Pi_d}{r} = \Pi^R \tag{5.16}$$

It is assumed that the profits that the incumbent makes when his patent is not challenged or is challenged and upheld (Π^u) are greater than the profits that he makes when his patent is challenged and revoked (Π^R), that is, $\Pi^u > \Pi^R$. This assumption guarantees that the incumbent is not indifferent between receiving and not receiving patent protection for his process; the incumbent is better off when he receives patent protection.

Substitution of equations (5.15) and (5.16) into equation (5.14) yields the expression for the incumbent's discounted expected profits when the patent is challenged:

$$E(\Pi_I)^C = \mu\Pi^u + (1-\mu)\Pi^R - C_T \tag{5.17}$$

Substitution of equations (5.15) and (5.17) into equation (5.13) yields the expression for the incumbent's discounted profits when entry is not deterred. Recall that the probability of the patent being challenged is $\delta = b$ and the probability of the patent being found valid is $\mu = 1-b$.

$$\begin{aligned} E(\Pi_I)^{ND} &= \delta\{\mu\Pi^u + (1-\mu)\Pi^R - C_T\} + (1-\delta)\Pi^u = (\delta\mu + 1 - \delta)\Pi^u + \\ &(\delta - \delta\mu)\Pi^R - \delta C_T = (1-b^2)\Pi^u + b^2\Pi^R - bC_T \end{aligned} \tag{5.18}$$

The analysis so far has proceeded assuming that the entrant will always find it optimal to enter the market. It has been shown that the incumbent cannot deter entry when the patent is challenged and revoked since in this case the entrant's profits upon entry do not depend on the incumbent's patent breadth (see equation 5.1). Recall that the exogenous sunk costs (F) were assumed to be such as to allow a duopolistic market structure. It has also been shown that when the patent is not challenged, or is challenged and upheld, there may exist a patent breadth value $\hat{b} \in (0,1]$ such that the entry condition is not satisfied, that is, $\Pi_d \le \dfrac{x^*(\hat{b},r,\Pi_d) + (r + \lambda(x^*(\hat{b},r,\Pi_d),\hat{b}))F}{\lambda(x^*(\hat{b},r,\Pi_d),\hat{b})}$. If $\hat{b}$ exists and it is chosen by the incumbent, the entrant will not enter when the patent is not challenged or is challenged and upheld and the incumbent will make monopoly profits. The incumbent's profits when the patent is not challenged or is challenged and upheld and the incumbent chooses patent breadth $\hat{b}$ are given by equation (5.19).

$$(\Pi_I)^{NC} = (\Pi_I)_U^C = \int_0^\infty e^{-rt}\Pi_m dt = \frac{\Pi_m}{r} \tag{5.19}$$

The incumbent's discounted expected profits when patent breadth $\hat{b}$ that deters entry is chosen are given by substituting equations (5.14), (5.19) and (5.16) into equation (5.13). The incumbent's discounted expected profits when entry is deterred are given by equation (5.20).

$$E(\Pi_I)^D = \delta\{\mu\frac{\Pi_m}{r} + (1-\mu)\Pi^R - C_T\} + (1-\delta)\frac{\Pi_m}{r} = \\ (1-\hat{b}^2)\frac{\Pi_m}{r} + \hat{b}^2\Pi^R - \hat{b}C_T \tag{5.20}$$

It should be noted that if a $\hat{b}$ that deters entry exists it will be chosen by the incumbent *if and only if* the incumbent's expected discounted profits when $\hat{b}$ is chosen are greater than or equal to his profits when entry is not deterred, $E(\Pi_I)^D \geq E(\Pi_I)^{ND}$. Thus, it may not always be optimal for the incumbent to deter entry in this model. To keep the model simple, the analysis proceeds assuming that either there is no patent breadth $\hat{b}$ that can deter entry or that if a patent breadth $\hat{b}$ exists, it is not optimal for the incumbent to deter entry because $\hat{b}$ does not satisfy the condition $E(\Pi_I)^D \geq E(\Pi_I)^{ND}$.

Given the assumption that entry will not be deterred the incumbent chooses the patent breadth that maximizes the expected discounted profits given by equation (5.18). His objective function is given by:

$$\max_b E(\Pi_I)^{ND} = (1-b^2)\Pi^u + b^2\Pi^R - bC_T \tag{5.21}$$

Optimization of equation (5.21) yields the F.O.C. for a maximum. The F.O.C. are given by equation (5.22).

$$\frac{\partial E(\Pi_I)^{ND}}{\partial b} = 0 \Rightarrow (-2b)\Pi^u + (1-b^2)\frac{\partial \Pi^u}{\partial b} + 2b\Pi^R - C_T = 0 \Rightarrow \\ b^* = b(\Pi_m, \Pi_d, C_T, r) \tag{5.22}$$

The F.O.C. for the incumbent's optimization problem yield the optimal choice of patent breadth as a function of known parameters; the monopoly profits, the duopoly profits, the legal costs of the challenge process and the discount rate.

The interpretation of the F.O.C. given in equation (5.22) requires the determination of the sign of the term $\frac{\partial \Pi^u}{\partial b}$. The term $\frac{\partial \Pi^u}{\partial b}$ shows how the expected profits made by the incumbent when his patent is not challenged or when it is challenged and upheld are affected by the breadth of patent protection. The affect of patent breadth on Π^u does not depend on the nature of the instantaneous probability of success, as shown in the next proposition.

***Proposition** 5.7 The expected profits made by the incumbent when his patent is not challenged or when it is challenged and upheld are increasing in patent breadth* ($\frac{\partial \Pi^u}{\partial b} > 0$) *for both the additive and the multiplicative formulations of the instantaneous probability of success,* λ.

Proof:

It is straight forward to prove that $\frac{\partial \Pi^u}{\partial b} > 0$ when the instantaneous probability of success is multiplicatively separable in the flow of R&D spending and in the patent breadth. In this case, an increase in patent breadth leads to a decrease in the flow of R&D spending, $x_b < 0$ (see Proposition 5.2). The term $\frac{\partial \Pi^u}{\partial b}$ is equal to $\frac{\partial \Pi^u}{\partial b} = \frac{(\lambda_b + \lambda_x x_b)(r\Pi_d - \Pi_m)}{(r+\lambda)^2}$, where the term $(r\Pi_d - \Pi_m)$ is negative as duopoly profits are always smaller than monopoly profits and where $\lambda_b < 0$ and $\lambda_x > 0$ from the theoretical assumptions made about the instantaneous probability of success. The above conditions imply that $\frac{\partial \Pi^u}{\partial b} > 0$. When the instantaneous probability of success is additively separable in the flow of R&D spending and in the patent breadth, an increase in patent breadth leads to an increase in the flow of R&D spending, $x_b > 0$ (see Proposition 5.1). In this case, given that $r\Pi_d - \Pi_m < 0$, $\lambda_b < 0$ and $\lambda_x > 0$, the sign of the term $\frac{\partial \Pi^u}{\partial b}$ depends on the sign of the expression $(\lambda_b + \lambda_x x_b)$. To determine the sign of the term $\frac{\partial \Pi^u}{\partial b}$, the additively separable function $f_1 : \lambda = x^\theta + \frac{1}{b}$ is used. Using f_1 the expression for $\frac{\partial \Pi^u}{\partial b}$ is given by $\frac{\partial \Pi^u}{\partial b} = \frac{\Pi_m - r\Pi_d}{(1 + b(r + x^\theta))^2}$ which is greater than zero $\forall \theta \in (0,1)$, $b \in (0,1]$, $x \geq 0$ and $r \in [0,1]$. □

The intuition behind the result presented in Proposition 5.7 is as follows. When the instantaneous probability of success is multiplicatively separable in the flow of R&D spending and in the patent breadth, an increase in patent breadth affects the instantaneous probability of success both directly ($\lambda_b < 0$) and indirectly ($\lambda_{xb} < 0$). Since the entrant responds to an increase in patent breadth with a decrease in the flow of R&D spending ($x_b < 0$), it becomes more difficult for the entrant to succeed in generating the non-infringing process. The more difficult it is for the entrant to succeed, the longer the incumbent can operate as a monopolist and the greater are his expected profits (Π^u). When the instantaneous probability of success is additively separable in the flow of R&D spending and in the patent breadth, an increase in patent breadth affects the instantaneous probability of success only directly ($\lambda_b < 0$ and $\lambda_{xb} = 0$). In addition, the entrant responds to an increase in patent breadth with an increase in the flow of R&D spending ($x_b > 0$). The increase in the flow of the R&D spending, in turn, has a positive affect on the instantaneous probability of success ($\lambda_x > 0$). The total effect of an increase in patent breadth on the incumbent's expected profits (Π^u) is positive because the decrease in the probability of success caused by an increase in patent breadth is greater than the increase in the probability of success caused by the increase in the flow of R&D spending (i.e., $\lambda_b > \lambda_x$).

Having determined how patent breadth affects the incumbent's expected profits when his patent is not challenged or when it is challenged and upheld ($\frac{\partial \Pi^u}{\partial b}$), the F.O.C. can be interpreted. The F.O.C. demonstrate the trade off that the incumbent faces when he determines the optimal breadth of patent protection. An increase in patent breadth leads to an increase in the incumbent's expected returns by $(1-b^2)\frac{\partial \Pi^u}{\partial b} + 2b\Pi^R$; this increase represents the marginal benefit to the incumbent from an increase in patent breadth. At the same time, an increase in patent breadth leads to a decrease in the incumbent's expected returns by $2b\Pi^u + C_T$; this decrease represents the marginal cost to the incumbent from an increase in patent breadth. Given that as patent breadth increases so does the probability that the patent will be challenged and revoked, by increasing patent breadth the incumbent increases the likelihood that he will realize profits Π^R (i.e., profits earned when the patent is revoked) rather than Π^u (i.e., profits earned when the patent is not challenged or is challenged and upheld). In addition, by increasing patent breadth the incumbent increases the profits made when the patent is not challenged or is challenged and upheld (since $\frac{\partial \Pi^u}{\partial b} > 0$) but, at the same time, he increases the probability that the patent will be challenged and that he will have to incur the legal costs C_T. At the optimal patent breadth the marginal benefits will be equal to the marginal costs.

To graphically characterize the determination of the optimal patent breadth let $k(b) = C_T$ and $f(b) = (-2b)\Pi^u + (1-b^2)\frac{\partial \Pi^u}{\partial b} + 2b\Pi^R$.[9] The F.O.C. for a maximum can then be written as follows:

$$f(b) - k(b) = 0 \Rightarrow (-2b)\Pi^u + (1-b^2)\frac{\partial \Pi^u}{\partial b} + 2b\Pi^R = C_T \tag{5.23}$$

The S.O.C. for a maximum imply that the following inequality must be satisfied

$$f_b - k_b < 0 \Rightarrow f_b < k_b \tag{5.24}$$

Given that the $k_b = 0$, the S.O.C. imply that $f_b < 0$ which means that f_b must cut k_b from above at the optimum. It is easily verified that f_b is decreasing in b as $f(b \to 0) \approx \frac{\partial \Pi^u}{\partial b} > 0$ while $f(b=1) = -2\Pi^u + 2\Pi^R < 0$. To guarantee the existence of an optimum the increase in the incumbent's expected profits when the patent is not challenged or challenged and upheld, $\frac{\partial \Pi^u}{\partial b}$, should be greater than the legal costs

[9] Note that the functions $k(b)$ and $f(b)$ are not defined in terms of marginal benefits and marginal costs because the slope and the curvature of the marginal benefit curve cannot be determined without knowledge of the values of the parameters that affect it. The chosen formulation simplifies the analysis.

incurred by the incumbent when the patent is challenged, C_T. The requirement that $\frac{\partial \Pi^u}{\partial b} > C_T$ guarantees that f_b cuts k_b from above.

Proposition 5.8 *Claiming the maximum breadth of patent protection (i.e., $b^*=1$) is never an optimal strategy for the incumbent in this model.*

Proof:

At b=1 $k(b=1) = C_T \geq 0$ and $f(b=1) = 2(-\Pi^u + \Pi^R) < 0$. The above imply that the curves $k(b)$ and $f(b)$ will never cross at b=1. The same result is of course derived when the marginal benefits and the marginal costs are compared for b=1. When b=1 the marginal costs are always greater than the marginal benefits, i.e., $2\Pi^u + C_T > 2\Pi^R$. Thus, b=1 is not a profit maximizing patent breadth choice for the incumbent in this model. □

The graphical representation of the determination of the optimal patent breadth is depicted in Figure 5.4. In Figure 5.4 the slope of the curve $f(b)$ has been assumed to be decreasing in patent breadth, $f_{bb} < 0$.[10]

[10] The curvature of *f*(b) cannot be determined without knowledge of the magnitude of the parameters that affect it. Note that the determination of the curvature of *f*(b) in not important for the results, it is necessary only for the graphical representation of the optimum.

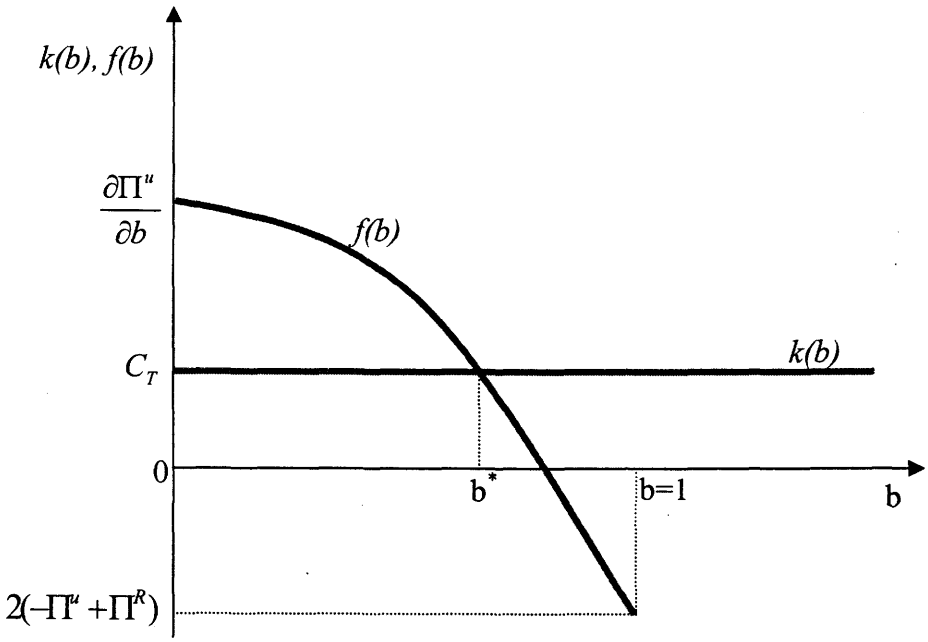


Figure 5.4 Graphical Representation of the Determination of the Optimal Patent Breadth

As shown in equation (5.22) the optimal patent breadth is a function of the following parameters, $b^* = (\Pi_m, \Pi_d, C_T, r)$. The effect of a change in the parameters of interest on the optimal patent breadth choice is determined by the signs of the following terms, $\frac{db^*}{d\Pi_m}, \frac{db^*}{d\Pi_d}$ $\frac{db^*}{dC_T}$ and $\frac{db^*}{dr}$.

The effect of a change in the monopoly profits on the optimal breadth is determined first. The expression for $\frac{db^*}{d\Pi_m}$ is derived by totally differentiating the optimality condition $f(b) - k(b) = 0 \Rightarrow (-2b)\Pi^u + (1-b^2)\frac{\partial \Pi^u}{\partial b} + 2b\Pi^R - C_T = 0$ with respect to the optimal patent breadth, b*, and the monopoly profits, Π_m:

$$\frac{\partial f(b^*)}{\partial b^*}db^* + \frac{\partial f(b^*)}{\partial \Pi_m}d\Pi_m = \frac{\partial k(b^*)}{\partial b^*}db^* + \frac{\partial k(b^*)}{\partial \Pi_m}d\Pi_m \Rightarrow$$
$$(f_b - k_b)db^* = (k_{\Pi_m} - f_{\Pi_m})d\Pi_m \Rightarrow \frac{db^*}{d\Pi_m} = \frac{k_{\Pi_m} - f_{\Pi_m}}{(f_b - k_b)} \tag{5.25}$$

From the S.O.C., $f_b - k_b < 0$. Since $k_{\Pi_m} = 0$ and $f_{\Pi_m} = -\frac{2b}{(r+\lambda)} - \frac{(1-b^2)(\lambda_b + \lambda_x x_b)}{(r+\lambda)^2}$,

the sign of the term $\frac{db^*}{d\Pi_m}$ depends on the sign of the term f_{Π_m}.

Proposition 5.9 *An increase in the monopoly profits leads to an increase in the optimal patent breadth* ($\frac{db^*}{d\Pi_m} > 0$), *when* $b^* \in (0,\bar{b})$ *and to a decrease in the optimal patent breadth* ($\frac{db^*}{d\Pi_m} < 0$), *when* $b^* \in (\bar{b},1]$. *The patent breadth* $\bar{b} \in (0,1]$ *is the breadth of patent protection that makes the effect of a change in monopoly profits on the optimal patent breadth equal to zero,* $\frac{d(b^* = \bar{b})}{d\Pi_m} = 0$. *The patent breadth* $\bar{b}$ *exists for both the additive and multiplicative formulations of the instantaneous probability of success.*

Proof:

The additively and multiplicatively separable functions $f_1 : \lambda = x^\theta + \frac{1}{b}$ and $f_2 : \lambda = \frac{x^\theta}{b}$ are used, respectively, to prove Proposition 5.9. The detailed proof is presented in the Appendix. It is found that there exists a patent breadth $\bar{b} \in (0,1]$ such that $f_{\Pi_m}(\bar{b}) = 0$ for both f_1 and f_2. It is also found that f_{Π_m} is decreasing in patent breadth $\forall\theta \in (0,1)$, $x \geq 0$ and $r \in [0,1]$. The above imply that if patent breadth b^* is such that $b^* \in (0,\bar{b})$ then

$f_{\Pi_m} > 0$ which implies that $\frac{db^*}{d\Pi_m} > 0$ while if patent breadth b* is such that $b^* \in (\bar{b},1]$ then $f_{\Pi_m} < 0$ which implies that $\frac{db^*}{d\Pi_m} < 0$ (see equation (5.25)). □

The intuition behind the results of Proposition 5.9 is as follows. There are two countervailing effects that take place as patent breadth increases. On the one hand, a larger patent breadth makes it harder for the entrant to succeed in generating a non-infringing process, thus allowing the incumbent to make monopoly profits for a longer period. On the other hand, the probability that the patent will be challenged and the probability that it will be revoked increase, making it less likely for the incumbent to realize monopoly profits. There is a critical patent breadth value $\bar{b}$ which makes the two effects equal. When the breadth of patent protection is smaller than $\bar{b}$, the danger of having the patent challenged and revoked is relatively small and the incumbent tries to capture the (increased) monopoly profits by making it harder for the entrant to succeed. Thus, when $b^*<\bar{b}$, an increase in the anticipated monopoly profits results in an increase in the optimal breadth of patent protection. However, when initially the patent breadth is greater than $\bar{b}$, the risk of having the patent revoked (due to the large patent breadth) is now relatively large and the incumbent reduces the breadth of protection in order to reduce the probability that the patent will be revoked and that he will not have a chance to operate as a monopolist. Thus, when initially patent breadth is greater than $\bar{b}$, the incumbent responds to an increase in the anticipated monopoly profits with a decrease in the optimal patent breadth level.

The effect of a change in the duopoly profits on the optimal breadth of patent protection claimed, $\frac{db^*}{d\Pi_d}$, is determined by totally differentiating the optimality condition $f(b)-k(b)=0 \Rightarrow (-2b)\Pi^u+(1-b^2)\frac{\partial \Pi^u}{\partial b}+2b\Pi^R - C_T=0$ with respect to the optimal patent breadth, b*, and the duopoly profits, Π_d. The expression for $\frac{db^*}{d\Pi_d}$ is given by equation (5.26).

$$
\begin{aligned}
&\frac{\partial f(b^*)}{\partial b^*}db^* + \frac{\partial f(b^*)}{\partial \Pi_d}d\Pi_d = \frac{\partial k(b^*)}{\partial b^*}db^* + \frac{\partial k(b^*)}{\partial \Pi_d}d\Pi_d \Rightarrow \\
&(f_b-k_b)db^* = (k_{\Pi_d}-f_{\Pi_d})d\Pi_d \Rightarrow \frac{db^*}{d\Pi_d} = \frac{k_{\Pi_d}-f_{\Pi_d}}{(f_b-k_b)}
\end{aligned} \tag{5.26}
$$

From the S.O.C. the term $f_b-k_b<0$. In addition, $k_{\Pi_d}=0$ and $f_{\Pi_d}=-\frac{2b\lambda}{(r+\lambda)}+\frac{r(1-b^2)(\lambda_b+\lambda_x x_b)}{(r+\lambda)^2}+\frac{2b}{r}$ which implies that the sign of the term $\frac{db^*}{d\Pi_d}$ depends on the sign of the term f_{Π_d}. Given the above, if $f_{\Pi_d}>0$ then $\frac{db^*}{d\Pi_d}>0$ while if $f_{\Pi_d}\leq 0$ then $\frac{db^*}{d\Pi_d}<0$.

Proposition 5.10 *An increase in the duopoly profits leads to a decrease in the optimal patent breadth ($\frac{db^*}{d\Pi_d}<0$), when $b^*\in(0,\overline{\overline{b}})$ and to an increase in the optimal patent breadth ($\frac{db^*}{d\Pi_d}>0$), when $b^*\in(\overline{\overline{b}},1]$. The patent breadth $\overline{\overline{b}}\in(0,1]$ is the breadth of patent protection that makes the effect of a change in duopoly profits on the optimal* *patent breadth equal to zero,* $\frac{d(b^* = \overline{\overline{b}})}{d\Pi_d} = 0$. *The patent breadth* $\overline{\overline{b}}$ *exists for both the additive and multiplicative formulations of the instantaneous probability of success.*

Proof:

The additively and multiplicatively separable functions $f_1 : \lambda = x^\theta + \frac{1}{b}$ and $f_2 : \lambda = \frac{x^\theta}{b}$ are used, respectively, to prove the above proposition. The detailed proof is presented in the Appendix. It is found that there exists a patent breadth $\overline{\overline{b}} \in (0,1]$ such that $f_{\Pi_d}(b^* = \overline{\overline{b}}) = 0$ for both f_1 and f_2. It is also found that the term f_{Π_d} is increasing in patent breadth, $\frac{\partial f_{\Pi_d}}{\partial b} > 0$, $\forall\theta \in (0,1)$, $x \geq 0$ and $r \in [0,1]$. The above imply that if patent breadth b^* is such that $b^* \in (0, \overline{\overline{b}})$ then $f_{\Pi_d} < 0$ which implies that $\frac{db^*}{d\Pi_d} < 0$ and if patent breadth b^* is such that $b^* \in (\overline{\overline{b}}, 1]$ then $f_{\Pi_d} > 0$ which implies that $\frac{db^*}{d\Pi_d} > 0$ (see equation (5.26)). □

The intuition behind the results of Proposition 5.10 is as follows. As discussed above two countervailing effects take place as patent breadth increases. On the one hand, it becomes harder for the entrant to succeed and on the other hand the probability that the patent will be challenged and the probability that it will be revoked increase. If the patent breadth is such that $b \in (0, \overline{\overline{b}})$, then the incumbent responds to an increase in duopoly profits by decreasing patent breadth to make it easier for the entrant to succeed and so that he can realize the duopoly profits. If the patent breadth is such that $b \in (\overline{\overline{b}}, 1]$, then the incumbent increases patent breadth to make it easier for his patent to be challenged and revoked, thus again increasing the probability of realizing the increased duopoly profits.

The effect of a change in the legal costs incurred by the incumbent on the optimal level of patent breadth is determined by totally differentiating the optimality condition $f(b)-k(b)=0 \Rightarrow (-2b)\Pi^{u}+(1-b^{2})\frac{\partial \Pi^{u}}{\partial b}+2b\Pi^{R}-C_{T}=0$ with respect to the optimal patent breadth, b*, and the legal costs C_T. The expression for $\frac{db^{*}}{dC_{T}}$ is given by:

$$\frac{\partial f(b^{*})}{\partial b^{*}}db^{*}+\frac{\partial f(b^{*})}{\partial C_{T}}dC_{T}=\frac{\partial k(b^{*})}{\partial b^{*}}db^{*}+\frac{\partial k(b^{*})}{\partial C_{T}}db^{*}\Rightarrow$$
$$(f_{b}-k_{b})db^{*}=(k_{C_{T}}-f_{C_{T}})dC_{T}\Rightarrow\frac{db^{*}}{dC_{T}}=\frac{k_{C_{T}}-f_{C_{T}}}{(f_{b}-k_{b})} \tag{5.27}$$

Proposition 5.11 *The effect of a change in the legal costs incurred by the incumbent on the optimal patent breadth is negative,* $\frac{db^{*}}{dC_{T}}<0$, *for both the additive and the multiplicative formulations of the instantaneous probability of success,* λ.

Proof:

From the S.O.C. the term $f_b-k_b<0$. In addition, $k_{C_T}=1$ while $f_{C_T}=0$ which imply that $\frac{db^{*}}{dC_{T}}<0$. □

The results of Proposition 5.11 are as expected. The more expensive it becomes for the incumbent to defend the patent during a patent validity challenge, the less willing is the incumbent to risk having the patent challenged. The incumbent decreases the probability of having the patent challenged by decreasing the breadth of patent protection.

Finally, the effect of a change in the discount rate on the optimal patent breadth is determined by totally differentiating the optimality condition $f(b)-k(b)=0 \Rightarrow (-2b)\Pi^{u}+(1-b^{2})\frac{\partial \Pi^{u}}{\partial b}+2b\Pi^{R}-C_{T}=0$ with respect to the optimal patent breadth, b^{*}, and the discount rate, r. The expression for $\frac{db^{*}}{dr}$ is given by:

$$\begin{aligned}&\frac{\partial f(b^{*})}{\partial b^{*}}db^{*}+\frac{\partial f(b^{*})}{\partial r}dr=\frac{\partial k(b^{*})}{\partial b^{*}}db^{*}+\frac{\partial k(b^{*})}{\partial r}dr \Rightarrow \\ &(f_{b}-k_{b})db^{*}=(k_{r}-f_{r})dr \Rightarrow \frac{db^{*}}{dr}=\frac{k_{r}-f_{r}}{(f_{b}-k_{b})}\end{aligned} \tag{5.28}$$

From the S.O.C. the term $f_{b}-k_{b}<0$. In addition, $k_{r}=0$ while

$$\begin{aligned}f_{r}=&\frac{2(-1+b^{2})(\lambda_{b}+\lambda_{x}x_{b})(r\Pi_{d}-\Pi_{m})}{(r+\lambda)^{3}}-\frac{2b\Pi_{d}}{r^{2}}+\frac{2b(\lambda\Pi_{d}+\Pi_{m})}{(r+\lambda)^{2}}+ \\ &\frac{(1-b^{2})(\lambda_{b}+\lambda_{x}x_{b})\Pi_{d}}{(r+\lambda)^{2}}\end{aligned}$$

The sign of the term f_{r} cannot be determined without knowledge of the magnitude of the parameters that affect it and thus the sign of the term $\frac{db^{*}}{dr}$ is inconclusive.

To summarize the main findings of sub-section 5.3.3, the incumbent's optimal patent breadth choice depends on the level of monopoly profits that the incumbent realizes for as long as the entrant does not succeed in generating a non-infringing process, the level of duopoly profits realized by the incumbent once the entrant succeeds, the legal costs incurred during the patent challenge process and the discount rate. Claiming the maximum breadth of patent protection (b_{max}=1) is not a profit maximizing strategy for the entrant in this model. The effect of a change in the level of monopoly and duopoly profits on the optimal patent breadth depends on the initial optimal patent breadth value. The effect of a change in the legal costs incurred by the incumbent during the patent challenge process on the optimal patent breadth choice is always negative while the effect of a change in the discount rate on the optimal patent breadth choice is inconclusive.

5.4 Concluding Remarks

The chapter uses a simple game theoretic model to model and to examine the determination of the optimal patent breadth for the innovator of a drastic process innovation. The optimal patent breadth for the innovator is the breadth of patent protection that maximizes the innovator's ability to appropriate innovation rents. The game consists of two players, an incumbent innovator who having generated a drastic process innovation and having decided to patent it determines the breadth of patent protection and an entrant who decides how much to spend on R&D to generate her own process.

The innovator in this model acts strategically and with foresight. That is, the innovator takes into consideration the entrant's response to his choice of patent breadth and the possibility that he may have to defend the validity of his patent when he determines the optimal breadth of patent protection claimed. The model allows for the probability that the patent will be challenged by a third party as soon as the patent is granted. The probability that the patent will be challenged and the probability that the validity of the patent will be upheld depend on the breadth of patent protection. The possibility of patent infringement is not considered in this model. It is thus assumed that if the entrant enters, she will do so without infringing the patent.

In this model, the R&D process is stochastic and the instantaneous probability of success is either additively or multiplicatively separable in the entrant's flow of R&D spending and in the incumbent's patent breadth choice. It is assumed that when success is realized by the entrant, her process is as efficient as the incumbent's process in producing the non-patentable product. Both players use their processes for the production of a new non-patentable product which is viewed as a homogenous product by consumers.

The results show that when the patent is revoked the entrant enters the market using the incumbent's process. When the patent is not challenged or is challenged and upheld, the entrant's optimal flow of R&D spending depends on the breadth of patent protection, the duopoly profits that the entrant will realize upon success and the discount rate. The effect of patent breadth on the entrant's optimal flow of R&D spending is positive or negative depending on whether the instantaneous probability of success is additively or multiplicatively separable, respectively, on the flow of R&D spending and on patent breadth. The duopoly profits have a positive effect on the optimal flow of R&D spending while the effect of the discount rate on the optimal flow of R&D spending is inconclusive.

The optimal breadth of patent protection depends on the level of monopoly profits realized by the incumbent during the period that the entrant undertakes R&D, the level of duopoly profits realized once the entrant succeeds, the legal costs incurred by the incumbent during the patent challenge process and the discount rate. The effect of the monopoly and the duopoly profits on the optimal patent breadth choice depends on the initial patent breadth value. The incumbent's legal costs have a negative effect on the optimal patent breadth while the effect of the discount rate on the optimal patent breadth is inconclusive.

The results show that there may exist a patent breadth that deters entry, but it may not be optimal for the incumbent to choose this patent breadth and deter entry. The results also show that claiming the maximum breadth of patent protection ($b_{max}=1$) is never an optimal strategy for the incumbent in this model. This result is not surprising, however, since the model assumes that when the maximum patent breadth is claimed the patent is always challenged and is revoked. Thus, since when the patent is revoked the entrant always enters using the incumbent's process, the incumbent never finds it optimal to claim the maximum patent breadth.

The results hold under the assumption of no patent infringement which implies that patent breadth affects the entrant's probability of success. If infringement was an option for the entrant then if the entrant found it optimal to infringe the patent, patent breadth would not have a binding effect on the entrant's probability of generating an infringing process. The results also depend on the assumption that the patent validity is challenged only by a third party. If the model allowed for a validity challenge by the entrant, as well as by a third party, then the optimal patent breadth might have been narrower. In addition, it has been assumed that there is only one entrant, that the patent life is infinite and that entry deterrence is either not possible or is not an optimal choice for the incumbent. Relaxing the above assumptions is the focus of future research.

CHAPTER VI

TOWARDS AN EMPIRICAL MODEL OF OPTIMAL PATENT BREADTH

6.1 Introduction

The theoretical models developed in chapters IV and V examined the determination of the optimal patent breadth for drastic product and process innovations, respectively. The theoretical models determined the variables that influence the innovator's patent breadth decision and described the effect of these variables on the optimal patent breadth.

The purpose of this chapter is to use the theoretical findings of chapters IV and V to develop an econometric model that could be used to empirically test the validity of the theoretical results/propositions. The theoretical findings of the two strategic patent breadth models are combined in this chapter to derive a single empirical model.

The rest of the chapter is organized as follows. Section 6.2 explains how the results of the theoretical models are used to derive the empirical model and an estimation process for the econometric model developed is proposed. In section 6.3 the selection of the patent sample that satisfies the theoretical assumptions and should thus be used in the empirical analysis is explored. Possible ways for empirically approximating the variables of interest are examined in section 6.3; this section also looks at possible data sources that could be used to obtain data for the variables of the model. Section 6.4 concludes the chapter.

6.2 Empirical Model Development

The theoretical models developed in chapters IV and V assumed that the patent breadth decision is made under complete and perfect information. This assumption implies that when the patentee decides on the breadth of patent protection claimed he can foresee who will be the potential competitors, the potential competitors' R&D effectiveness and trial costs, the level of *anticipated* monopoly and/or duopoly profits, as well as his own trial costs that will be incurred if the patent is infringed or challenged. The assumption of perfect and complete information is satisfied by assuming perfect foresight in the development of the empirical model.

The theoretical models suggest that, in determining patent breadth, the patentee takes into consideration a number of variables. These variables are the potential competitors' R&D effectiveness (β) and legal costs incurred during an infringement trial (C_E), the patentee's legal costs incurred during an infringement trial (C_P) and during a direct validity challenge (C_T), the level of monopoly profits (Π_m) realized by the patentee for as long as entry does not occur, the level of duopoly profits (Π_d) realized by the patentee and the entrant when entry occurs, and the discount rate (r).

As described in Proposition 4.10, the individual effect of the entrant's R&D effectiveness (β) on patent breadth is positive.[1] The individual effect of the entrant's trial costs (C_E) on patent breadth could not be determined in the theoretical models without additional knowledge about the values of certain parameters. The infringement trial (C_P) and validity challenge (C_T) legal costs incurred by the patentee have an expected negative effect on patent breadth as described in Propositions 4.10 and 5.11.[2] The expected effect of the monopoly (Π_m) and duopoly (Π_d) profits on patent breadth could be negative or positive depending on the patent breadth value as described in Propositions 5.5 and 5.9. Finally, the effect that the discount rate (r) has on patent breadth could not be determined in the theoretical models without additional knowledge of the values of certain parameters.

The above variables are all exogenous; their values do not depend on the choice of patent breadth. In addition to the exogenous variables, the determination of patent breadth depends on a number of endogenous variables. These endogenous variables are the competitors' R&D spending ($x = f(b, \Pi_d, r)$), the probability that the patent will be challenged (δ=b) and/or infringed (see Proposition 4.2 below), as well as the probability that the patent will be upheld (μ=1-b) and the probability ($\lambda = f(x, b)$) that the entrant will succeed in generating a competing innovation.

---

[1]The greater are the entrant's R&D costs, the greater is the patentee's incentive to induce infringement, which implies that the greater is the chosen patent breadth. This is so because according to Proposition 4.2, the greater is patent breadth, the greater is the entrant's incentive to infringe the patent.

[2] According to Proposition 4.10, the greater are the patentee's trial costs, the smaller is the patentee's incentive to induce infringement and thus, the smaller must be the chosen patent breadth. This last result occurs because, as shown in Proposition 4.2, patent breadth is positively related to the entrant's incentive to infringe the patent.

These endogenous variables affect the determination of optimal patent breadth because, along with the exogenous variables β and C_E, they determine whether the legal costs (C_P and C_T) will be incurred by the patentee and whether the patentee will realize monopoly (Π_m) or duopoly profits (Π_d). Recall that the infringement trial (C_P) and validity challenge (C_T) legal costs are incurred only when the patent is infringed and the patentee files an infringement suit and/or when the validity of the patent is directly challenged. In addition, the monopoly profits are realized by the patentee only when entry in the market can be blocked and/or when infringement is found during trial and the infringer is not allowed in the market and/or until the entrant succeeds in generating a competing innovation. The duopoly profits are realized by the patentee when the patentee cannot deter entry and the competitor succeeds in generating a competing process and/or when the patent is invalidated or infringement is not found.

In addition, according to the theoretical models, an important determinant of the optimal patent breadth choice is the *combination* of the entrant's R&D effectiveness (β) and trial cost (C_E) values. The combination of the β and C_E values determines whether there is a patent breadth, $\hat{b}$, that can deter entry (i.e., $\frac{8}{9\beta} \leq \hat{b} \leq \sqrt{\frac{81C_E\beta}{8}}$) and/or whether there is a patent breadth, $\tilde{b}$, that can make the entrant indifferent between infringing and not infringing the patent (i.e., $\tilde{b} = \frac{9(4\beta + \sqrt{2}\sqrt{\beta}\sqrt{16C_E + 8\beta + 81C_E\beta^2})}{16 + 81\beta^2}$). These combinations are depicted in Figure 4.8 in chapter IV. The following propositions summarize the main results of the two theoretical models that will be used in the development of the empirical model.

Proposition 4.2 *When the entrant finds it optimal to enter the market (i.e., when entry cannot be deterred) then:*

(a) The greater is the breadth of patent protection the greater is the entrant's incentive to infringe the patent.

(b) The more costly it is to produce the better quality product the greater is the entrant's incentive to infringe the patent.

(c) The greater are the entrant's trial costs the smaller is the entrant's incentive to infringe the patent.

Proposition 4.10 *When the patentee cannot deter entry (a $\hat{b}$ does not exist) and there exists a patent breadth $\tilde{b}$ that makes the entrant indifferent between infringing and not infringing the patent then:*

(a) The greater are the patentee's monopoly profits (Π_m) the greater is the patentee's incentive to induce infringement.

(b) The greater are the patentee's trial costs (C_P) the smaller is the patentee's incentive to induce infringement.

(c) The greater are the entrant's costs of producing the better quality product the greater is the patentee's incentive to induce infringement given that the patentee's monopoly profits are different than zero.

Proposition 5.1 *When the instantaneous probability of success, λ, is additively separable in patent breadth, b, and in the flow of R&D spending, x, (i.e., $\lambda = \varphi(x) + \psi(b)$) the effect of a change in patent breadth on the optimal flow of R&D spending is positive, i.e.,* $\frac{dx^*}{db} > 0$.

Proposition 5.2 *When the instantaneous probability of success, λ, is multiplicatively separable in the patent breadth, b, and in the flow of R&D spending, x, (i.e., $\lambda = \varphi(x) \cdot \psi(b)$) the effect of a change in patent breadth on the optimal flow of R&D spending is negative, i.e.,* $\frac{dx^*}{db} < 0$.

Proposition 5.5 *The optimal level of the flow of R&D spending is increasing in the duopoly profits that the entrant anticipates to make if she succeeds in generating a non-infringing process,* $\frac{dx^*}{d\Pi_d} > 0$, *for both the additive and multiplicative formulations of the instantaneous probability of success, λ.*

Proposition 5.9 *An increase in the monopoly profits leads to an increase in the optimal patent breadth* ($\frac{db^*}{d\Pi_m} > 0$), *when* $b^* \in (0,\bar{b})$ *and to a decrease in the optimal patent* *breadth* ($\frac{db^*}{d\Pi_m} < 0$)*, when* $b^* \in (\bar{b}, 1]$. *The patent breadth* $\bar{b} \in (0,1]$ *is the breadth of patent protection that makes the effect of a change in monopoly profits on the optimal patent breadth equal to zero,* $\frac{d(b^* = \bar{b})}{d\Pi_m} = 0$. *The patent breadth* $\bar{b}$ *exists for both the additive and multiplicative formulations of the instantaneous probability of success.*

Proposition 5.10 *An increase in the duopoly profits leads to a decrease in the optimal patent breadth* ($\frac{db^*}{d\Pi_d} < 0$)*, when* $b^* \in (0, \bar{\bar{b}})$ *and to an increase in the optimal patent breadth* ($\frac{db^*}{d\Pi_d} > 0$)*, when* $b^* \in (\bar{\bar{b}}, 1]$. *The patent breadth* $\bar{\bar{b}} \in (0,1]$ *is the breadth of patent protection that makes the effect of a change in duopoly profits on the optimal patent breadth equal to zero,* $\frac{d(b^* = \bar{\bar{b}})}{d\Pi_d} = 0$. *The patent breadth* $\bar{\bar{b}}$ *exists for both the additive and multiplicative formulations of the instantaneous probability of success.*

Proposition 5.11 The *effect of a change in the legal costs incurred by the incumbent on the optimal patent breadth is negative,* $\frac{db^*}{dC_T} < 0$, *for both the additive and multiplicative formulations of the instantaneous probability of success,* λ.

The above results will be used for the development of the empirical model. The econometric model should account for the endogeneity that is present in the theoretical models. The following three binary (0/1) variables are defined for this purpose.

$$y_1 = \begin{cases} 1, \text{ if the patent is infringed} \\ 0, \text{ otherwise} \end{cases} \qquad y_2 = \begin{cases} 1, \text{ if the patent is challenged} \\ 0, \text{ otherwise} \end{cases}$$

$$y_3 = \begin{cases} 1, \text{ if entry occurs} \\ 0, \text{ otherwise} \end{cases}$$

Given Propositions 4.2 and 4.10 the variable y_1 (patent infringement), is a function of the variables β, C_E, b, Π_m and C_P. Specifically, $y_1 = f(\underset{+}{\beta}, \underset{-}{C_E}, \underset{+}{b}, \underset{+}{\Pi_m}, \underset{-}{C_P})$. The signs beneath the variables indicate the direction of the effect that these variables have on variable y_1 as described in Propositions 4.2 and 4.10. Thus, the more effective is the R&D process (the smaller is β), the smaller is the incentive to infringe the patent. The greater are the entrant's (C_E) and the patentee's (C_P) trial costs, the smaller is the incentive to infringe the patent and induce infringement, respectively. The greater is the breadth of the patent (b), the greater is the entrant's incentive to infringe it while the greater are the anticipated monopoly profits (Π_m), the greater is the patentee's incentive to induce infringement when entry cannot be deterred.

The variable y_2 (patent challenge) is a function of the variables β, C_E and b. Specifically, $y_2 = f(\underset{+}{\beta}, \underset{-}{C_E}, \underset{+}{b})$. This specification reflects the different processes at work in the determination of optimal patent breadth in the product innovation and process innovation cases. Chapter V (which looks at process innovations) explicitly models the decision to challenge the validity of the patent. In the model of chapter V it is assumed that the probability that the patent will be challenged depends only on patent breadth because of the assumption that a third party and not the potential competitor challenges the validity of the patent. Chapter IV (which looks at product innovations) implicitly models the decision to challenge the validity of the patent. The model of chapter IV assumes that once the entrant locates within the patentee's claims a trial takes place either because the incumbent files an infringement lawsuit or because the validity of the patent is directly challenged. Thus, in the model of chapter IV the probability that the patent will be challenged (which is equivalent to whether the entrant will locate within the patentee's claims) depends on the breadth of the patent but also on the entrant's R&D effectiveness (β) and trial costs (C_E). The greater are the entrant's R&D costs (the greater is β) and the greater is patent breadth, the greater is the incentive to challenge the patent. The greater are the entrant's trial costs, the smaller is the incentive to challenge the patent.

The variable y_3 (entry occurrence) depends on the variables βC_E, b and x. Specifically, $y_3 = f(\underset{\pm}{\beta C_E}, \underset{\pm}{b}, \underset{+}{x})$. Entry occurs when the entrant's R&D effectiveness and trial costs (i.e., the combination of β and C_E values) are such that entry cannot be deterred by the patentee, when the patent is revoked and/or when the entrant succeeds in generating a competing innovation. When the combination of β and C_E values is such that entry can be deterred (i.e., $\beta \geq \frac{8}{9}$ and $C_E \geq \frac{512}{6561\beta^3}$), the effect of $C_E\beta$ on y_3 is negative. When the combination of β and C_E values is such that entry cannot be deterred (i.e., $\frac{4}{9} \leq \beta < \frac{8}{9}$ and $C_E > \frac{16-72\beta+81\beta^2}{162\beta}$ or $C_E \leq \frac{16-72\beta+81\beta^2}{162\beta}$ and

$C_E < \frac{512}{6561\beta^3}$), the effect of $C_E\beta$ on y_3 is positive. Patent breadth (b) affects y_3 in two different ways. On the one hand, the greater is patent breadth, the greater is the probability (1-μ=b) that the patent will be revoked and thus, the greater is the likelihood that the entrant will enter. On the other hand, the greater is patent breadth the harder it is for the entrant to succeed in developing a competing innovation and thus, the smaller is the likelihood that the entrant will enter. The total effect of patent breadth on y_3 depends on the relative magnitude of the above two effects. The effect of the flow of R&D spending (x) on y_3 is positive because the greater is x, the greater is the probability (λ) that success will be realized and thus, the greater is the probability that the entrant will enter.

The following structural form of the model captures the simultaneous relationships outlined above:

$$b = a_0 + a_1 y_1 + a_2 y_2 + a_3 y_3 + a_4 \beta + a_5 C_E + a_6(\beta C_E) + a_7 C_P + a_8 C_T + a_9 \Pi_m + a_{10}\Pi_d + a_{11} r + u_1 \tag{6.1}$$

$$y_1 = \Phi(c_0 + c_1\beta + c_2 C_E + c_3 b + c_4 \Pi_m + c_5 C_P) + u_2 \tag{6.2}$$

$$y_2 = \Phi(d_0 + d_1\beta + d_2 C_E + d_3 b) + u_3 \tag{6.3}$$

$$y_3 = \Phi(e_0 + e_1(\beta C_E) + e_2 b + e_3 x) + u_4 \tag{6.4}$$

$$x = g_0 + g_1 b + g_2 \Pi_d + g_3 r + u_5 \tag{6.5}$$

Equations (6.1) shows that when the patentee determines the optimal breadth of patent protection he takes into account the exogenous variables β (the entrant's R&D effectiveness), C_E (the entrant's trial costs), r (the discount rate), Π_m (the monopoly profits), Π_d (the duopoly profits) and C_T (his own trial costs). He also takes in to account the endogenous variables that determine the probability of occurrence of some of the exogenous variables (i.e., y_1 and y_2 determine whether C_T and C_P will be incurred, while y_3 determines whether Π_m or Π_d will be realized). The term βC_E gives the joint effect of the entrant's R&D effectiveness and trial costs on the patent breadth choice; note that it is the combination of the β and C_E values that determines whether there is a patent breadth that deters entry and/or a patent breadth that can make the entrant indifferent between infringing and not infringing the patent.

Equations (6.2), (6.3) and (6.4) are probit equations where the function $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and the u_i's are $u_i \sim$ *iid* N $(0, \sigma^2)$, i=2,3,4. The relationship between the discrete choice dependent variables y_1, y_2 and y_3 and the variables that influence them was described earlier.

Finally, equation (6.5) shows how the entrant determines her optimal R&D spending. Given Propositions 5.1 and 5.2, the effect that patent breadth has on R&D spending is positive or negative depending on whether the instantaneous probability of success is additively or multiplicatively separable in x and b, respectively (i.e., whether patent breadth affects the marginal product of R&D spending or not). Proposition 5.5 shows that the effect of the duopoly profits on the R&D spending is positive. The effect that the discount rate has on the R&D spending could not be determined in the theoretical models without knowledge of the value of certain parameters.

The above econometric model is not a conventional simultaneous equations model as some of the structural equations are non-linear probability equations. Given the non-linearity of equations (6.2), (6.3) and (6.4), it is extremely difficult to solve the structural equations for the endogenous variables to derive the reduced form of the model which is traditionally used to retrieve the structural parameters. In fact, there is no assurance that the reduced form of the model exists in this case (Greene 1997). In addition, even if the reduced form parameters were determinable (i.e., equations (6.2), (6.3) and (6.4) were not highly non-linear), the structural parameters of equation (6.1) could not be retrieved since this equation is underidentified.

A different estimation strategy is required to circumvent the estimation problems of the structural model. The estimation process that is developed involves two stages. In the first stage, the endogenous variables y_1, y_2, y_3 and x are regressed on all the exogenous variables of the system, i.e., β, C_E, βC_E, C_T, C_P, Π_m, Π_d and r. In the second stage, the fitted values of the endogenous variables are substituted into equation (6.1) and patent breadth is estimated.

More specifically, the estimation process takes place as follows. In the first stage, equation (6.6) is estimated first and the fitted value of x, $\hat{x}$, is obtained.

$$x = g_0' + g_1'\beta + g_2'C_E + g_3'\beta C_E + g_4'C_T + g_5'C_P + g_6'\Pi_m + g_7'\Pi_d + g_8'r + v_4 \tag{6.6}$$

The fitted value of x, $\hat{x}$, is substituted into equation (6.9) – see below. Then, the probit equations (6.7), (6.8) and (6.9) are separately estimated and the fitted probability values of the variables y_1, y_2, and y_3, are obtained, $\hat{y}_1$, $\hat{y}_2$ and $\hat{y}_3$.

$$y_1 = \Phi(c_0' + c_1'\beta + c_2'C_E + c_3'\beta C_E + c_4'C_T + c_5'C_P + c_6'\Pi_m + c_7'\Pi_d + c_8'r) + v_1 \tag{6.7}$$

$$y_2 = \Phi(d_0' + d_1'\beta + d_2'C_E + d_3'\beta C_E + d_4'C_T + d_5'C_P + d_6'\Pi_m + d_7'\Pi_d + d_8'r) + v_2 \tag{6.8}$$

$$y_3 = \Phi(e_0' + e_1'\beta + e_2'C_E + e_3'\beta C_E + e_4'C_T + e_5'C_P + e_6'\Pi_m + e_7'\Pi_d + e_8'r + e_9'\hat{x}) + v_3 \tag{6.9}$$

The last stage in the estimation process involves substitution of the fitted values $\hat{y}_1, \hat{y}_2$ and $\hat{y}_3$ into equation (6.1). The substitution yields equation (6.10), which can then be used to empirically estimate the determinants of the breadth of patent protection.

$$\begin{aligned} b = {} & a_0' + a_1'\hat{y}_1 + a_2'\hat{y}_2 + a_3'\hat{y}_3 + a_4'\beta + a_5'C_E + a_6'(\beta C_E) + a_7'C_P + a_8'C_T + \\ & a_9'\Pi_m + a_{10}'\Pi_d + a_{11}'r + u_1 \end{aligned} \tag{6.10}$$

The estimation of equation (6.10) will allows a determination of the effect that the right hand side variables have on observed patent breadth for the patent sample chosen and for the period of study. The above estimation process is an application of the Two-Stage Least Squares estimation process which is known to yield consistent estimates (Greene 1997). Thus, the estimation process described above leads to consistent estimates of the parameters of the variables in equation (6.10).

This section described the development of an empirical model that could be used to empirically estimate how certain variables affect the determination of the breadth of patent protection. The empirical model was developed using the theoretical results of the models developed in chapters IV and V. The estimation process that should be followed to estimate the empirical model was also described in this section.

The next section describes the process that should be followed to select the patent sample of interest (i.e., patents that satisfy the assumption of the theoretical models) from the entire patent population. Section 6.3 also describes how the variables of the empirical model given by equations (6.6) – (6.10) could be empirically measured and identifies possible data sources for the variables of interest.

6.3 Patent Sample Selection, Variable Approximation and Data Sources

The biotechnology and the pharmaceutical industries will be considered for the selection of the patent sample that will be used in the empirical analysis. As will become evident below, patents in these two industries are more likely to satisfy the assumptions made in the theoretical models. The patent sample could include patents issued by the USPTO and the EPO whose owners are North-American firms. This selection is made because of the availability of various data sources that document the financial and operating performance of North-American firms.

The time period considered for the empirical study could cover the period starting at the early 1970s (when the biotechnology industry was born) through to the late 1990s. The selection of the patents in the empirical analysis should stop at the late 1990s to allow for the possibility of a direct validity challenge and/or infringement of the patents in the sample. Note that direct validity challenges usually occur within a year from patent grant (Harhoff and Reitzig 2000), while the majority of patents are imitated within the first four years after grant (Mansfield et al. 1981).

As mentioned above, the patents that should be examined in the empirical analysis have to satisfy certain conditions/assumptions that are present in the theoretical models. The main assumptions of the theoretical models are that the innovator has no help from the Patent Office in determining the optimal patent breadth, that the patent protects a drastic innovation (product or process) and that the market in which the patentee operates is concentrated. Thus, to empirically examine the patenting behavior of innovators and to compare the observed behavior to that predicted by the theory, only patents that satisfy the above conditions should be considered.

The assumption that the patentee has no help from the Patent Office in structuring his claims implies that only innovations that were granted as filed by the Patent Office should be considered for the empirical analysis. The selection of patents that were granted as filed by the Patent Office could also enable the identification of patents that protect drastic innovations. This is so because when drastic innovations are concerned the patent examiner cannot find support from prior art (since there is no relevant prior art) to object to the claims (see discussion in chapter II).

The backward citations made in the patent and the forward citations that the patent receives can also be used as indicators of whether the innovation protected by the patent is drastic. Both the above measures have been used in empirical studies to approximate the value of the patent (Trajenberg 1990, Harhoff and Reitzig 2000, Lanjouw and Schankerman 2001). The justification for using the backward and forward citations as indicators of whether the innovation protected by the patent is drastic is as follows. A patentee must cite all relevant prior art in the patent specification, since failure to cite relevant prior art is sufficient grounds for invalidating the patent (Merges 1997). Thus, the number of backward citations in a patent is an indicator of whether the patented innovation breaks new ground (small number of citations due to limited or non-existent prior art) or whether it builds on previous innovations (large number of citations). The number of forward citations reflects the innovation's contribution to the development of subsequent innovations; the greater the number of forward citations a patent receives the more path-breaking or drastic the innovation is considered to be (Lanjouw and Schankerman 2001).

To determine the number of forward and backward citations that would be used to separate drastic from non-drastic innovations, the range and the average number of forward and backward citations for all biotechnology and pharmaceutical patents granted during the period of study should be determined.[3] The above statistics could be used to determine which number of forward citations is considered to be large and which number of backward citations is considered to be small for patents in the industries under study.

Another important assumption of the theoretical models is that the markets considered are concentrated and can support only a small number of products/processes. Given the above assumptions, patents protecting innovations in industries that are concentrated should be selected. Good candidates for the empirical analysis are patents protecting innovations in the pharmaceutical industry and in the biotechnology industry. Biotechnology and pharmaceutical patents can be identified on the basis of the International Patent Classification (IPC) assignments (Harhoff and Reitzig 2000).

The empirical model developed in section 6.2 assumed explicitly that the market structure will be either a monopoly or a duopoly. This is a very restrictive assumption and it is likely to be difficult to obtain a sufficient sample of patents that were granted as filed, that protect drastic innovations and that possess this degree of concentration. The above problem can be circumvented by allowing the patentee to face more than one competitor. In this case, in the empirical model, the duopoly profits (Π_d) should be approximated by the oligopoly profits made by the patentee when/if more than one competitor enter his market during the period of study.

---

[3] Harhoff and Reitzig (2000), in a study that examined what exposes patents to the risk of a direct validity challenge using a sample of 13,389 European biotechnology and pharmaceutical patents, found that the number of forward citations received by a patent ranged from 0 to 36 with an average of 1 citation per patent and the number of backward citations made by a patent ranged from 0 to 29 with an average of 2 citations per patent.

Given the above, the patents used in the empirical analysis should have the following characteristics: they should protect biotechnology and pharmaceutical innovations, they should have been granted as filed by the Patent Office, and they should have a 'large' number of forward citations and a 'small' number of backwards citations. Data on the above mentioned patent characteristics can be obtained from the following sources. Patent data on patents granted by the USPTO can be obtained from the USPTO's two CD-ROM databases (these databases are marketed as the BIB and ASSIST files of the CASSIS database) and by the BRS Information Technologies' PATDATA database (Lerner 1995). Patent data on patents granted by the EPO can be obtained by the ELPAC and the ESPACE databanks (Harhoff and Reitzig 2000). The above databases provide information on the inventor and the patent owner (if different), the designated states, IPC assignments, all information concerning the filing and granting procedures (i.e., number of claims, forward and backward citations for every patent, whether the patent was granted as filed) as well as information on oppositions decisions (direct validity challenges to the Patent Office).

Having selected the patent sample that should be considered in the empirical analysis, the next step in the estimation of the empirical model is the determination of the empirical measurement of the variables in the model. Patent breadth has been measured in empirical studies either by the number of claims in a patent application or by the number of classes and subclasses in the IPC system assigned to the patent by the patent examiner (Lanjouw and Schankerman 2001, Harhoff and Reitzig 2000, Lerner 1994). The number of classes and subclasses in the IPC system reflect the different technological areas to which the innovation is relevant. In general, the greater is the number of independent patent claims, the greater is the breadth of the patent. Likewise, the greater is the number of classes and subclasses assigned to the patent, the greater is the breadth of the patent.

The appropriate measure of patent breadth, b, to be used in the empirical model is the number of claims in the patent application. This is so because the theoretical models examine the determination of the patent breadth *claimed* by the patent applicant. The number of classes and subclasses in the IPC system, however, reflects the breadth of protection *assigned* to the patent by the patent examiner and not the breadth of protection claimed by the innovator.

The variables y_1 (patent infringement), y_2 (patent challenge), y_3 (entry occurrence), x (R&D spending), and r (discount rate) are all observable. Data on litigated/infringed patents can be obtained by the Patent History CD-ROM produced by Derwent (Lanjouw and Schankerman 2001). Lerner (1995) sites seven different data sources for obtaining data on litigated patents. Among these sources are the LIT-REEX data in LEXIS's PATENT/ALL database, the LIT-ALERT data provided by Research Publications Inc., and the LEXIS's PATENT/FEDCTS database which contains information on all previous litigation involving any given firm. Data on direct validity challenges can be obtained by the USPTO's and the EPO's patent databases and from the ELPAC, ESPACE and REFI databanks (Harhoff and Reitzig 2000). The above databases should be searched using the patent number and the name of the firm.

Data on the firms that are active in any given industry can be obtained by the Venture Economics' Venture Intelligence Database which assigns a four-digit industry classification number to each firm (e.g., 4000 class for biotechnology) and prepares a profile of the company's activities (Lerner 1995). The Recombinant Capital Database that specializes in collecting information about biotechnology firms could also be used to identify operating firms when the biotechnology industry is examined (Lerner 1995). Entry in any given market, as well as the number of firms/competitors, can be identified using the above databases.

Data on R&D expenditure for US firms can be obtained from the Industry Research and Development Information System (IRIS) database. This database provides R&D expenditure data on all published industry R&D data since 1953. The data on industry R&D expenditures is broken down by industry, size of company, size of R&D program, type of cost (e.g., wages or materials), source of funds, and as percent of company sales. Data on R&D expenditure for Canadian firms can be obtained from the Canadian Corporate R&D database. This database contains information on the research activities of 330 Canadian companies.

The reported discount rate, r, usually refers to the rates charged by Reserve Banks when they extend credit to depository institutions. The real discount rate at the time the patent application was filed should be used in the empirical analysis because this discount rate would reflect the innovator's evaluation of anticipated returns and costs at the time the patent breadth decision was made. Data on the market discount rates at different time periods can be obtained from any Federal Reserve Bank and these rates can be adjusted for inflation to obtain the real discount rates.

The variable β that refers to the rival firms' R&D effectiveness is not directly observable. The Research on Research (ROR) Committee initiated an industrial R&D database project with the Center for Innovation Management Studies (CIMS) where a Technology Value Program (TVP) was developed to measure the effectiveness of the R&D process. Some of the metrics used by the TVP to assess R&D effectiveness included new product sales, cost savings, patents at the firm and business-segment level and the projected values of the R&D pipeline (Larson 1999, Bean et al. 2000). Given the above, the variable β could be empirically measured by the ratio of the output of the R&D process over R&D expenditures. The output of the R&D process could be approximated by the number of patents granted to the company. This is an adequate proxy when industries that rely heavily on patent protection as a means of appropriating innovation rents are examined, like the biotechnology and the pharmaceutical industry.[4]

The variables Π_m (monopoly profits) and Π_d (duopoly/oligopoly profits) could be empirically measured as follows. The earnings of the given product line made by the firm for as long as it was alone in the market could be used to approximate monopoly profits while the earnings of the given product line made by the firm after entry (if entry occurred) would account for oligopoly profits.[5] Various data sources can be combined to get data on firms' earnings. When biotechnology patents are considered, data on earnings can be obtained by the Recombinant Capital Database. Earnings data on public firms can be received from the U.S. Securities and Exchange Commission (SEC) Database. State records,[6] the Venture Economics database, the COMPUSTAT database and the Global Vantage database can be used to obtain earnings data on private firms.

The variables C_E and C_P that denote litigation costs for the entrant and the incumbent, respectively, and the variable C_T, that denotes legal costs incurred by the patentee during a direct validity challenge in the Patent Office, are not directly observed.

---

[4] Trade secrecy is not a good alternative to patent protection for the biotechnology industry. As Lerner (1995) points out, the lack of reliance on trade secrecy is due to the strong academic roots of the industry and to the high mobility of its human capital. If industries where only a fraction of the innovations are patented were considered, the output of the R&D process could be approximated by the number of new products introduced by the company.

[5] Earnings equal revenues minus costs of sales, operating expenses, and taxes.

[6] As Lerner (1995) points out, private firms in many states have to make detailed filings reporting their financial and operating performance. These filings are publicly available.

Lerner (1995) used two proxies to empirically measure litigation costs; the number of previous patent suits in which the firm has been involved and the firm's financial resources measured by the paid-in capital (i.e., capital received from investors for stocks). The greater the number of lawsuits in which the firm was previously involved, the smaller are its litigation costs. According to Lerner (1995) this is so because 'there is a substantial "learning curve" in patent litigation, as firms become more efficient in managing internal and external counsel' (Lerner 1995, p. 475). Lerner (1995) uses the firm's paid-in capital as a measure of the size of new biotechnology firms. The justification for using the size of the firm as a proxy for litigation costs is that the larger is the firm, the more able it is to employ internal corporate patent counsel and thus the less costly is litigation for the firm.[7]

The number of previous patent lawsuits in which the firm has been involved can be used as a proxy for litigation costs in this model. Following Lerner (1995), firms can be divided into different categories based on the number of previous lawsuits; zero previous lawsuits, one to five previous lawsuits, six to ten previous lawsuits and so on. As mentioned above, the LEXIS's PATENT/FEDCTS database can be used to obtain information on all previous litigation involving any given firm.

Paid-in capital is not an adequate proxy for litigation costs in this model as the present study does not distinguish between new and established firms. As Lerner (1995) points out, paid-in capital is a less meaningful measure of the size of the firm when

---

[7] Lerner (1995) cites previous studies that indicate that smaller firms have higher litigation costs (i.e., Pashigian 1982).

established firms are concerned.[8] Instead, the market valuation of the firm could be used as a proxy of the size of the firm and of the firm's litigation costs. Firms can be divided into different categories based on their market valuation; low, medium and high, where the medium may refer to the average of the sample under consideration. Data on the firms' market valuation could be obtained from the SEC filings, the Venture Economics, the Recombinant Capital and the Global Vantage databases (Lerner 1995).

A similar method could be used to approximate the legal costs incurred by the patentee during an opposition process (direct validity attacks at the Patent Office). These costs are denoted by the variable C_T. Thus, the number of previous validity challenges in which the firm has been involved and its market valuation could be used to empirically approximate the legal costs incurred by the patentee during the direct patent validity challenge. Data on the number of previous validity challenges in which the firm has been involved, both as a challenger and as a challenged party, could be obtained from the Patent Offices' databases mentioned earlier.

Having described the measures that could be used to empirically approximate the variables in the model and having identified possible data sources for the variables of interest, the econometric model given by equations (6.6) – (6.10) could be empirically estimated. The resulting empirical estimates will depict the effect that the variables of interest have on the patentee's patent breadth decision for patents protecting drastic pharmaceutical and biotechnology innovations over the period of study. The empirical results could then be compared to the predictions made by the theoretical models to

---

[8] Established firms are more likely to have raised external equity in their earlier stages of their development and they may appear to have modest paid-in capital even though they may have substantial financial resources (Lerner 1995).

examine whether the observed/estimated patenting behavior of patentees in the biotechnology and pharmaceutical industry resembles the strategic patenting behavior described by the theoretical models.

6.4 Concluding Remarks

The results of the theoretical models developed in chapters IV and V were combined in this chapter to enable the development of an empirical model. The empirical model developed can be used to analyze the observed patenting behavior of innovators in specific industries during a given period of study. The estimated patenting behavior can then be compared to the strategic patenting behavior that is proposed by the theoretical models as an optimal way to appropriate innovation rents.

The main findings of the theoretical models presented in the form of propositions and the construction of three binary variables enabled the development of the empirical model. The variables y_1 (patent infringement), y_2 (patent challenge), and y_3 (entry occurrence) were introduced to capture the endogeneity that is present in the determination of the optimal patent breadth decision.

The empirical model developed is a simultaneous equations model in which some of the equations are non-linear probability equations. The non-linearity of some of the equations in the model is unlikely to allow the derivation of the reduced form of the model, the technique that is traditionally used to derive the structural parameters. A two step process is developed instead to enable the estimation of the simultaneous equations model. The estimation process is an application of the Two-Stage Least Squares estimation process which is known to yield consistent estimates.

The patent sample selected for the empirical estimation should satisfy the main assumption of the theoretical models; the patent should protect a drastic innovation, the patentee cannot rely on the Patent Office for help in structuring his claims and the patentee operates in a concentrated market. Biotechnology and pharmaceutical patents with a relatively large number of forward and a small number of backward citations, that were granted as filed by the Patent Office, could be selected for the empirical analysis. The different data sources that could be used to select data on the variables of the model are described in this chapter.

The development of the theoretical models of chapters IV and V should be seen as a first step towards the study of the determination of the patent breadth that would maximize the innovator's ability to appropriate innovation rents. Although some of the assumptions present in the theoretical model are restrictive, their purpose is to simplify the theoretical analysis. Relaxing some of these restrictive assumptions will enable the development of a more realistic empirical model.

CHAPTER VII

SUMMARY AND CONCLUSIONS

7.1 Summary

Patent protection has been used as a mechanism for the generation and dissemination of technological knowledge. The provision of patent protection has contributed to the development and growth of research intensive and innovative industries, like the biotechnology industry, by enabling innovators to better capture the returns from their innovations. The degree of appropriability of innovation rents enabled by a patent is mainly determined by patent length and patent breadth. Patent length, the time period during which exclusive rights on the innovation are granted by the patent is predetermined by law and cannot be influenced by the innovator's decisions. Patent breadth defines the technological territory claimed and protected by the patent and is explicitly chosen by the innovator, determined by the Patent Office and refined by the Patent Office and/or the courts.

The innovator determines the breadth of patent protection sought through the claims made in the patent application. In general, the greater the number of claims made and the more general they are, the broader is the patent protection claimed. The innovator's patent breadth decision is of great importance since claims determine both the patentability of his/her innovation (claims must satisfy the patentability requirements of novelty, utility, nonobviousness and enablement) and the viability of his/her patent after the grant (claims are routinely challenged in the Patent Office and/or the courts).

Existing studies of the innovator's patenting behavior have focused on the analysis of the innovator's decision to either patent the innovation or to keep it a secret. The innovator's patenting behavior once the decision to patent has been made – his/her decision on the breadth of patent protection claimed – has not been explicitly modeled in the literature. Instead, various studies have implicitly or explicitly assumed that the innovator will attempt to maximize the appropriability of his/her innovation by claiming the broadest patent protection possible.

The choice of maximum patent breadth may appear – on the surface – to be a profit-maximizing strategy for the incumbent since, the greater is patent breadth, the harder it is for competitors to enter the incumbent's market and thus the longer the incumbent can enjoy the limited monopoly position. However, the broader is the patent protection claimed, the harder it is to secure a patent grant. This occurs because the greater is patent breadth, the harder it is to differentiate the innovation from prior art (i.e., to show novelty), to demonstrate that the innovation is not obvious and to provide an enabling disclosure. In addition, even if the Patent Office grants the broad claims, the broader is the patent protection, the greater is the probability that the patent will be infringed, that it will overlap another patent and that the Patent Office and/or the courts will invalidate it or narrow its scope.

Given the above, a profit-maximizing innovator needs to take into consideration both the marginal benefits and the marginal costs of an increase in patent breadth to determine the optimal breadth of patent protection claimed. The innovator needs to be particularly concerned about patent breadth when (s)he seeks patent protection for a drastic innovation. When drastic innovations are concerned, the innovator cannot depend as much on the Patent Office for help in structuring his/her claims. This occurs because the more drastic is the innovation, the harder it is for the patent examiner to find support in the relevant prior art (since there is no relevant prior art) to object to broad or erroneous claims. Drastic innovations are thus usually granted broader protection by the Patent Office. Broad patent protection, combined with the large innovation rents that are often associated with drastic innovations, result in patents that are more vulnerable to validity attacks.

The main objective of this study is to examine the optimal patent breadth strategy that an innovator should employ when faced with the possibility that the patent breadth claimed will be challenged. In achieving this objective, the study first examines the patent granting and the patent challenge processes, as well as the economic literature on patent breadth, to provide evidence of the economic importance of the patent breadth decision for the innovator's ability to capture innovation rents. Having demonstrated that patent breadth is a strategic variable for the innovator, the study then models the innovator's strategic patent breadth decision.

In specific, the study explicitly models the patenting behavior of an innovator who determines the optimal breadth of patent protection claimed for a drastic product and a drastic process innovation. Two different models are developed to study the innovator's patent breadth decision. The first model, developed in chapter IV, examines the patent breadth decision for a drastic product innovation, while the second model, developed in chapter V, examines the patent breadth decision for a drastic process innovation. The results of the above models are used to develop an empirical model that could be used to study the patenting behavior of innovators in various industries.

The study explores the innovator's patenting behavior under different assumptions with respect to the environment in which the innovator operates. Thus, the patent breadth model for drastic product innovations assumes that the products produced by the incumbent and the entrant are vertically differentiated, while the patent breadth model for drastic process innovations assumes that the processes developed by the incumbent and by the entrant are equally efficient in producing a given product, which is viewed as a homogeneous product by consumers. In addition, the former model assumes that the R&D process is deterministic and that patent infringement is a threat to the incumbent while the latter model allows for a stochastic R&D process and assumes that patent infringement is never a threat to the incumbent.

Both models assume that the innovator has no help from the Patent Office in structuring his claims. In both models the patent breadth decision is modeled as a sequential game of complete and perfect information between an incumbent/patentee and a potential entrant. The incumbent behaves strategically and with foresight, taking into consideration the effect that his claims have on his potential competitor and the possibility that he may have to defend or enforce his patent rights after the patent grant.

The model developed to study the innovator's strategic patent breadth decision for drastic product innovations determines the optimal patent breadth for an innovator who faces a positive probability of infringement and/or a validity challenge. The incumbent anticipates entry of only one entrant who produces a better quality product. The innovator moves first, deciding on the breadth of patent protection claimed, while the entrant moves next deciding whether to enter the market and, if entry occurs, where to locate in the quality product space (i.e., to infringe the patent or not). In the last stage of the game the players choose their respective prices and compete in the market.

In the drastic product innovation model, the entrant has an incentive to locate as far as possible from the incumbent (i.e., to maximize the degree of differentiation between her product and the incumbent's product), since this relaxes price competition in the last stage of the game. Generating a better quality product is increasingly costly for the entrant, however. The model predicts that for certain values of the entrant's trial costs and R&D effectiveness there exists a patent breadth that can deter entry. If such a patent breadth exists, then it is always optimal for the incumbent to choose this patent breadth and deter entry. The results also suggest that claiming the maximum patent breadth is not always a profit-maximizing strategy for the incumbent. Instead, only when it is never optimal for the entrant to infringe the patent given her trial costs and R&D effectiveness it is optimal for the incumbent to claim the maximum patent breadth.

The model developed to examine the innovator's patent breadth decision for drastic process innovations determines the optimal patent breadth for the innovator when he faces a positive probability of a direct validity challenge by a third party. The model assumes that there is only one entrant who threatens the incumbent's monopoly position. The incumbent moves first, determining the breadth of patent protection. The entrant moves next deciding how much to spend on R&D to develop her own non-infringing process. The entrant moves after observing whether the patent has been challenged and the outcome of the challenge. In this model, the entrant has an incentive to locate as close as possible to the patentee in the process space since the incumbent's process is the most efficient way of generating the product and the products produced by the incumbent's and the entrant's processes are viewed as homogenous by consumers. Thus, when the patent is revoked the entrant enters the market using the incumbent's process.

The model assumes that if the patent is not challenged, or is challenged and upheld, the entrant enters the market only if she succeeds in developing a non-infringing process.

The drastic process innovation model predicts that the optimal patent breadth is a function of the monopoly profits that the incumbent makes for as long as the entrant is not successful in generating her own non-infringing process, the duopoly profits realized once the entrant succeeds, the incumbent's trial costs and the discount rate. The results also show that there may exist a patent breadth that deters entry, but it may not be optimal for the incumbent to choose this patent breadth and deter entry. In addition, the model predicts that the incumbent will never find it optimal to claim the maximum breadth of patent protection. This result is not surprising, however, since the model is built on the assumption that when the maximum breadth of protection is claimed the patent is always challenged and revoked.

The results of the game theoretic models are also used to develop the framework for an empirical model that could be used, in a future study, to examine the patenting behavior of innovators in different industries. The present study describes the selection of the patent sample that satisfies the theoretical assumption and could thus be used in the empirical analysis, explores possible ways of approximating the variables of interest and identifies various data sources for the variables of the empirical model.

The analysis conducted in this dissertation sheds some important light on the innovator's patent breadth decision. Despite the stylized nature of the theoretical models, some useful insights were provided. The innovator's profit-maximizing patent breadth decision involves a trade off. On the one hand, a broad patent protection increases the innovator's expected short-run returns by making it harder for competitors to enter his/her market without infringing the patent. On the other hand, a broad patent protection puts the viability of the patent at risk by increasing the probability that the patent will be infringed, that its validity will be challenged and that the courts will invalidate the patent or narrows its scope. Generally this trade off implies that the innovator will choose a patent breadth that is less than the maximum possible. In addition, under certain circumstances, this less than maximum patent breadth will deter entry; however, deterring entry might not always be a profit-maximizing strategy for the innovator.

7.2 Relaxing the Assumptions

The analysis conducted in this study considers a particular institutional setting. The institutional setting is such that the innovator cannot look to the Patent Office for help in structuring his claims, a fencepost patent system is in place and only one entrant can potentially enter. Modifying the institutional setting is likely to change the results. The remainder of this section examines the likely impact of relaxing some of the assumptions that were made in the models developed in chapters IV and V. A more complete analytical examination of the issues discussed below is the focus of future study.

In the development of both theoretical models it was assumed that there is only one Patent Office where the incumbent could apply for a patent, or equivalently that all Patent Offices are the same in terms of the protection that they grant. Patent Offices differ with respect to the protection that they grant, however. For instance, the USPTO is believed to grant broader patents than the EPO (Merges and Nelson 1990). In addition, some patent systems favor the fencepost and others the signpost approach to claims interpretation.

When the assumption of a single Patent Office is relaxed, the innovator's decision in which Patent Office to seek patent protection might affect his/her decision on the patent breadth claimed. For instance, if patent protection is sought in a Patent Office that is known to usually grant broad patent protection, the innovator might be better off claiming a narrower patent breadth than if protection was sought in a Patent Office that does not usually grant broad patents. This might occur because, in the former case, the innovator (as was assumed in this study) cannot depend on the Patent Office for help in narrowing broad claims that might not survive a validity challenge. The magnitude of the effect of the innovator's decision in which Patent Office to apply for patent protection on his/her decision regarding the patent breadth claimed depends on how concerned the innovator is about having to defend and/or enforce his/her patent rights after the grant.

The assumption of a fencepost patent system made in both theoretical models is a valid assumption when patent protection is sought in the EPO, since, according to Cornish (1989), the fencepost approach has been preferred by the European patent system. When patent systems that favor the signpost approach are considered, however, a different model may be required to examine the innovator's patent breadth decision.

The assumption of a fencepost patent system implies that patent claims define an exact border of protection. The implications of the assumption of a fencepost system for the drastic product innovation model is that when the entrant locates within the incumbent's claims, infringement will always be found unless the patent is found to be invalid. The implication is that, under a fencepost system, the probability that the patent is found to be infringed (once the entrant locates within the incumbent's patent claims) and the probability that the patent is found to be invalid are mutually exclusive and exhaustive. In addition, the fencepost assumption implies that it is not important how close to the patentee the entrant locates for infringement to be found.

The implications of the assumption of a fencepost patent system for the drastic process innovation model is that since patent claims define an exact border of protection, the entrant can determine whether her process infringes the patent or not. That is, as long as the entrant locates outside of the incumbent's claims in the process space her process will not infringe the patent. Patent breadth under a fencepost patent system is thus an important determinant of, and has a binding effect on, the entrant's probability of generating a non-infringing process.

Relaxing the assumption of a fencepost patent system (i.e., assuming a signpost patent system instead) has a number of implications for each model. First, assuming a signpost system implies that patent infringement is not directly linked to patent validity. That is, there could be a case where, even though the entrant locates within the patentee's claims, infringement is not found and the validity of the patent is upheld at the same time. This occurs because, in such a case, infringement is not found not because the entrant proved that the patent is invalid but because, using the doctrine of reverse equivalents, the entrant's product/process was found to be substantially different from the incumbent's. In such a case, it is important how close to the patentee the entrant locates because, the closer the entrant locates to the patentee, the greater is the probability that infringement will be found (i.e., the harder it is to show non infringement using the doctrine of reverse equivalents). Second, when the assumption of a fencepost system is relaxed, infringement may be found even when the entrant locates outside the patentee's claims (using the doctrine of equivalents). Thus, the entrant cannot determine based on her location choice whether her product and/or her process will infringe the patent or not.

Relaxing the assumption of a fencepost patent system may change the behavior of both the incumbent and the entrant. Under a signpost system the probability that infringement will be found depends on both the breadth of patent protection (which affects the probability that the validity of the patent will be upheld) and on the distance away from the incumbent that the entrant locates (which affects the doctrines used to determine infringement). In addition, patent breadth does not have the same binding effect as under a fencepost system on the probability of success, in the sense that now the entrant is not constrained to locate outside the incumbent's claims to not infringe the patent and, at the same time, even if she locates outside the patentee's claims, infringement may be found. In other words, there is uncertainty for both the incumbent and the entrant with respect to what constitutes infringement. This uncertainty is likely to change the way both the incumbent and the entrant behave. Thus, the incumbent might claim a narrower patent protection under a signpost than under a fencepost system in order to increase the probability that infringement will be found when the entrant locates within his/her claims. This narrowing might occur because, the smaller is the patent breadth claimed, the closer the entrant must locate to the incumbent to locate within the incumbent's claims. However, the closer the entrant locates to the incumbent, the harder it becomes to prove non infringement using the doctrine of reverse equivalents. For the same reason, the entrant is more likely to infringe a broad rather than a narrow patent under a signpost system.

Another assumption that is important for the results of the two models is the assumption of a single entrant. Relaxing this assumption is expected to change the incumbent's optimal patent breadth choice. When more than one potential entrant is anticipated by the incumbent, the probability that (s)he will have to defend his/her patent or enforce his/her patent rights increases. In addition, the incumbent should be prepared to incur greater legal costs since (s)he may have to defend his/her patent against infringement more than once. The above may lead to a narrower patent protection claimed by the incumbent. The same effect on the incumbent's patent breadth choice could occur by relaxing the assumption that the patent is only challenged by a third party.

The optimal patent breadth decision under multiple entries will be influenced by the number of the potential competitors, their R&D costs, their legal costs and their strategies with respect to the incumbent and with respect to each other. Thus, when there is more than one potential competitor, the incumbent's patenting decision is more complex since (s)he needs to take into consideration issues like the outcome of a patent race between the potential entrants, whether the entrant who succeeds first will also seek patent protection and how the entrants that have not succeeded will react to the new patent and to the incumbent's patent (e.g., entrants that have not succeeded may decide to challenge the incumbent's patent when an entrant succeeds and receives patent protections, since there is less 'room' for them to enter the market).

As mentioned above, a narrower patent breadth claimed is more likely under multiple entries than under a single entry. The incumbent might still find it optimal to claim a broad patent protection under multiple entries, however. This might occur when the potential entrants' legal costs are high, in which case infringement and/or validity challenges by competitors are less probable. In addition, the incumbent may claim broad patent protection when his/her own legal costs are low and when the potential entrants' R&D costs are low (i.e., to make it harder for the entrants to succeed).

Finally, it would be interesting to examine how the results of the theoretical models developed in this study are affected by the assumptions made with respect to the nature of the probabilities of infringement, direct validity challenge and patent revocation. This study assumed that an infringement trial always takes place when the entrant locates within the incumbent's claims. Relaxing this assumption implies that the incumbent's decision whether to take the alleged infringer to court affects, and is affected by, the incumbent's optimal patent breadth decision. Besides patent breadth, the incumbent's decision to sue for infringement also depends on the incumbent's and the infringer's legal costs. In this case, the incumbent has to consider that claiming a broad patent protection might impede his/her ability to enforce his/her patent rights by taking the alleged infringer to court. This occurs because, the greater is patent breadth, the less likely it becomes that the patent will survive an indirect validity attack during the infringement trial. Thus, it is more likely that the incumbent will claim a narrow patent protection when (s)he is concerned about his/her patent being infringed.

In addition, it has been assumed that the probability that the patent will be challenged and that it will be revoked are only a function of patent breadth and that their relationship to patent breadth is linear. One would expect, however, that both the entrant's and the incumbent's legal costs may play an important role in determining the magnitude of these probabilities.[1] That is, the more the players spend on legal representation, the greater are their chances of winning at trial given that the more

[1] Note that in the model developed in chapter V the probability that the validity of the patent will be challenged is not affected by the entrant's trial costs because it is assumed that a third party and not the potential entrant challenges the validity of the patent.

'capable' attorneys charge higher fees. In addition, as Lerner (1995) points out, litigation can be less costly for large firms that can afford to employ internal corporate patent counsel which gives them an advantage over smaller firms.

Relaxing the assumption that the probabilities of a validity attack and patent invalidation depend only on patent breadth may result in broader patent protection claimed by incumbents who can afford legal representation of a higher quality and/or by incumbents with internal patent counsel. Relaxing the assumption that when the maximum breadth of patent protection is claimed the patent is always challenged and revoked may also result in broader patent protection claimed. In general, the smaller the effect of patent breadth on the above probabilities, the broader is expected to be the patent breadth claimed.

The framework of analysis developed in this dissertation is of particular importance for the agribusiness sector. Changing consumer preferences and the emerging industrialization of agriculture have made product differentiation and the development of cost reducing and quality enhancing processes important strategies for agribusiness firms. Agribusiness firms need to invest in R&D to be able to develop new products (create new markets), develop different versions of existing products (i.e., horizontally differentiated products), and/or develop processes that will reduce production costs (i.e., process innovations) and/or allow them to generate improved versions of existing products (i.e., vertically differentiated products). As agribusiness firms undertake R&D they have to consider whether to patent the outcome of their research. Firms will patent their innovations if patenting enables the firms to better capture the rents from their innovations and cover their research costs. Since it is the breadth of the patent protection that determines, to a large extent, the degree of appropriability of innovation rents enabled by the patent, the ability of innovating firms to determine the optimal scope of patent protection claimed is critical both for their economic performance and the development and growth of the agribusiness sector.

REFERENCES

Barton, J. H. "Intellectual Property Rights: Reforming the Patent System." *Science* 287(5460, March 2000): 1933-1939.

Bean, S. A., M. J. Russo and R. L. Whiteley. "Benchmarking Your R&D: Results from IRI/CIMS Annual R&D Survey for FY '98." Available at http://www.onlinejournal.net/iri/RTM/free/html/43_1_16.html, (2000).

Beathe, J. and Y. Katsoulacos. *The Economic Theory of Product Differentiation.* Cambridge University Press, 1991.

Boehlje, M. "Industrialization of Agriculture, What are the Implications?" *Choices* (First Quarter 1996): 30-33.

Canadian Patent Office (CPO). "Claims." *Manual of Patent Office Practice, Chapter 11.* Available on line at http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch11-e.pdf, (March 1998c).

Canadian Patent Office (CPO). "Classification." *Manual of Patent Office Practice, Chapter 12.* Available on line at http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch12-e.pdf, (March 1998d).

Canadian Patent Office (CPO). "Description." *Manual of Patent Office Practice, Chapter 9.* Available on line at http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch09-e.pdf, (March 1998b).

Canadian Patent Office (CPO). "Examination of Applications." *Manual of Patent Office*

Practice, Chapter 13. Available on line at http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch13-e.pdf, (March 1998e).

Canadian Patent Office (CPO). "Final Action Practice." *Manual of Patent Office Practice, Chapter 21.* Available on line at http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch21-e.pdf, (March 1998h).

Canadian Patent Office (CPO). "Opening and Inspection of Documents." *Manual of Patent Office Practice, Chapter 2.* Available on line at http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch02-e.pdf, (March 1998a).

Canadian Patent Office (CPO). "Requirements for Patentability." *Manual of Patent Office Practice, Chapter 15.* Available on http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch15-e.pdf, (March 1998f).

Canadian Patent Office (CPO). "Utility and Non-Statutory Subject Matter." *Manual of Patent Office Practice, Chapter 16.* Available on line at http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/mopop/ch16-e.pdf, (March 1998g).

Carlton, D. and J. Perloff. *Modern Industrial Organization.* Third Edition, Addison-Wesley Inc., 1999.

Chang, H.F. "Patent Scope, Antitrust Policy and Cumulative Innovation." *Rand Journal of Economics* 26(1, Spring 1995): 34-57.

Choi, P.J. "Patent Litigation as an Information-Transmission Mechanism." *The American Economic Review* 88(5, December 1998): 1249-1263.

Cornelli, F. and M. Schankerman. "Patent Renewals and R&D Incentives." *Rand*

Journal of Economics 30(2, Summer 1999): 197-213.

Cornish, W.R. *Intellectual Property: Patents, Copyright, Trademarks and Allied Rights*. New York: Matthew Bender & Company, 1989.

Cremer H. and J.F. Thisse. "Location Models of Horizontal Differentiation: A Special Case of Vertical Differentiation Models." *The Journal of Industrial Economics* 39(4, June 1991): 383-390.

Drabenstott, M. "Industrialization: Steady Current or Tidal Wave?" *Choices* (Fourth Quarter 1994): 4-8

Eswaran, M. and N. Gallini. "Patent Policy and the Direction of Technological Change." *Rand Journal of Economics* 27(4, Winter 1996): 722-746.

European Patent Office (EPO). "Guidelines for Examination in the European Patent Office." *Part C–Guidelines for Substantive Examination*. Available on line at, http://www.european-patent-office.org/legal/gui_lines/pdf/gui_e_full.pdf, (2000a).

European Patent Office (EPO). *Facts and Figures*. Available on line at http://www.european-patent-office.org/epo/an_rep/1999/pdf/ff99.pdf, (2000b).

Fudenberg, D. and J. Tirole. *Game Theory*. The MIT Press, Cambridge Massachusetts, 1991.

Gallini, N. "Patent Policy and Costly Imitation." *RAND Journal of Economics* 23(1, Spring 1992): 52-63.

Gallini, N. and B. Wright. "Technology Transfer under Asymmetric Information." *Rand Journal of Economics* 21(1, Spring 1990): 147-160.

Gilbert, R. and C. Shapiro. "Optimal Patent Length and Breadth." *RAND Journal of*

Economics 21(1, Spring 1990): 106-112.

Gilbert, R. and D. Newbery. "Preemptive Patenting and the Persistence of Monopoly." *The American Economic Review* 72(3, June 1982): 514-526.

Green, J. and S. Scotchmer. "On the Division of Profit in Sequential Innovation." *Rand Journal of Economics* 26(1, Spring 1995): 20-33.

Greene, W.H. *Econometric Analysis,* 3rd Edition, Macmillan, 1997.

Harhoff, D. and M. Reitzig. "Determinants of Opposition Against EPO Patent Grants – The Case of Biotechnology and Pharmaceuticals." Working Paper. Institute for Innovation Research and Technology Management, University of Munich. March 2000.

Hausman, J. and J. Mason. "Price Discrimination and Patent Policy." *Rand Journal of Economics* 19(2, Summer 1988): 253-265.

Hofer, M. and P. Edmundson. "Intellectual Property Insurance: A New Approach to Managing IP Litigation." *The Journal of Biolaw and Business* 3(1, 1999): 45-48.

Horstmann, I., G. M. MacDonald and A. Slivinski. "Patents as Information Transfer Mechanisms: To Patent or (Maybe) Not to Patent." *Journal of Political Economy* 93(5, October 1985): 837-858.

Jewkes, J., D. Sawers, and R. Stillerman. *The Sources of Invention*. London: Macmillan, 1960.

Kamien, M. and Y. Tauman. "Fees versus Royalties and the Private Value of a Patent." *Quarterly Journal of Economics* (August 1986): 471-491.

Klemperer, P. "How Broad Should the Scope of Patent Protection Be?" *RAND Journal of Economics* 21(1, Spring 1990): 113-130.

Lane, W.J. "Product Differentiation in a Market with Endogenous Sequential Entry." *RAND Journal of Economics* 11(1, Spring 1980): 237-260.

Lanjouw, J.O. and M. Schankerman. "Characteristics of Patent Litigation: A Window on Competition." *Rand Journal of Economics* 32(1, Spring 2001): 129-151.

Larson F. C. *Measuring the Value of the R&D in Industry*, GMU/NASA Workshop on Performance Metrics for R&D Organizations. Available at http://www.hq.nasa.gov/office/codez/plans/R&D/larson.pdf.

Lee, T. and L. Wilde. "Market Structure and Innovation: A Reformulation." *Quarterly Journal of Economics*, 94(2, March 1980): 429-436.

Lentz, E.T. "Adequacy of Disclosures of Biotechnology Inventions." *AIPLA Quarterly Journal* 16(1988): 314-338.

Lerner, J. "Patenting in the Shadow of Competitors." *Journal of Law and Economics* 38(2, October 1995): 463-495.

Lerner, J. "The Importance of Patent Scope: An Empirical Analysis." *Rand Journal of Economics* 25(2, Summer 1994): 319-333.

Levin, R., A. Klevorick, R. Nelson and S. Winter. "Appropriating the Returns from Industrial Research and Development." *Brookings Papers on Economic Activity* (3, 1987): 783-820.

Loury, G. "Market Structure and Innovation." *Quarterly Journal of Economics*, 93(3, August 1979): 395-410.

Machlup, F. *An Economic Review of the Patent System.* Study of the Subcommittee on Patents, Trademarks and Copyrights of the Committee on the Judiciary, United States Senate, Eighty-fifth Congress, Second Session, Study No. 15. Washington, D.C.: U.S. Government Printing Office, 1958.

Machlup, F. and E. Penrose. "The Patent Controversy in the Nineteenth Century." *The Journal of Economic History* 10(1, May 1950): 1-29.

Mansfield, E., M. Schwartz and S. Wagner. "Imitation Costs and Patents: An Empirical Study." *The Economic Journal* 91(December 1981): 907-918.

Matutes, C., P. Regibeau and K. Rockett. "Optimal Patent Design and the Diffusion of Innovations." *Rand Journal of Economics* 27(1, Spring 1996): 60-83.

McKeough, J. *Intellectual Property. Commentary and Materials.* The Law Book Company Limited, Second Edition, 1992.

Merges, R.P. and R. R. Nelson. "On the Complex Economics of Patent Scope." *Columbia Law Review* 90(4, May 1990): 839-916.

Merges, R.P. *Patent Law and Policy: Cases and Materials*, 2nd Edition, Charlottesville, Penn.: MICHIE Law Publishers, 1997.

Miller, A. and M. Davis. *Intellectual Property. Patents, Trademarks and Copyrights in a Nutshell.* West Publishing Company, Second Edition, 1990.

Motta, M. "Endogenous Quality Choice: Price vs. Quantity Competition." *The Journal of Industrial Economics* XLI(2, June 1993): 113-131.

Nordhaus, D.W. "The Economics of Patents." *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change.* The M.I.T. Press, 1969.

O'Donoghue, T. "A Patentability Requirement for Sequential Innovation." *Rand Journal of Economics* 29(4, Winter 1998): 654-679.

O'Donoghue, T., S. Scotchmer and J. Thisse. "Patent Breadth, Patent Life, and the Pace of Technological Progress." *Journal of Economics and Management Strategy*, 7(1, Spring 1998): 1-32.

Pashigian, P. "A Theory of Prevention and Legal Defense with an Application to the Legal Costs of Companies." *Journal of Law and Economics* 25(1982): 247-270.

Phillips, P. and D. Dierker. "Public Good and Private Greed: Strategies for Realizing Public Benefits from a Privatized Global Agrifood Research Effort." In P. G. Pardey, ed., *The Future of Food: Biotechnology Markets and Policies in an International Setting*, Washington D.C.: International Food Policy Research Institute, 2001.

Reinganum, J.F. "Uncertain Innovation and the Persistence of Monopoly." *The American Economic Review* 73(4, September 1983): 741-748.

Rockett, K. "Choosing the Competition and Patent Licensing." *Rand Journal of Economics* 21(1, Spring 1990): 161-171.

Schankerman, M. "How Valuable is Patent Protection? Estimates by Technology Field." *Rand Journal of Economics* 29(1, Spring 1998): 77-107.

Schankerman, M. and S. Scotchmer. "Damages and Injunctions in Protecting Intellectual Property." *Rand Journal of Economics* 32(1, Spring 2001): 199-220.

Scherer, F. M. "Nordhaus's Theory of Optimal Patent Life: A Geometric Reinterpretation." *The American Economic Review* 62(1972): 422-427.

Scherer, F. M. *Industrial Market Structure and Economic Performance.* Boston: Houghton Mifllin, 1980. Second edition.

Scotchmer, S. "On the Optimality of the Patent Renewal System." *Rand Journal of Economics* 32(2, Summer 1999): 181-196.

Scotchmer, S. "Protecting Early Innovations: Should Second-Generation Products be

Patentable?" *Rand Journal of Economics* 27(2, Summer 1996): 322-331.

Scotchmer, S. "Standing on the Shoulders of Giants: Cumulative Research and Patent Law." *Journal of Economic Perspectives* 5(1, Winter 1991): 29-41.

Scotchmer, S. and J. Green. "Novelty and Disclosure in Patent Law." *Rand Journal of Economics* 21(1, Spring 1990): 131-146.

Shaked, A. and J. Sutton. "Relaxing Price Competition through Product Differentiation." *The Review of Economic Studies* 49(1, January 1982): 3-13.

Shaked, A. and J. Sutton. "Natural Oligopolies." *Econometrica* 51(5, 1983): 1469-83.

Tandon, P. "Optimal Patents with Compulsory Licensing." *Journal of Political Economy* 90(3, 1982): 471-486.

Tirole, J. *The Theory of Industrial Organization.* The MIT Press, Cambridge Massachusetts, 1988.

Trajtenberg, M. "A Penny for your Quotes: Patent Citations and the Value of Innovations." *Rand Journal of Economics* 21(1, Spring 1990): 172-187.

U.S. Patent and Trademark Office (USPTO). *A Guide to Filling a Utility Patent Application.* Available on line at http://www.uspto.gov/web/offices/pac/utility/utility.htm, (March 2001a).

U.S. Patent and Trademark Office (USPTO). *Eighteen-month Publication of Patent Applications.* Available on line http://www.uspto.gov/web/offices/pac/pgpub.htm, (March 2001b).

U.S. Patent and Trademark Office (USPTO). *General Information Concerning Patents.* Available on line at http://www.uspto.gov/web/offices/pac/doc/general/index.html, (August 1999).

Van Dijk, T. "Patent Height and Competition in Product Improvements." *The Journal of Industrial Economics* XLIV(2, June 1996): 151-167.

Van Dijk, T. "The Economic Theory of Patents: A survey." *MERIT Research Memorandum 2/94-017*. Available on line at http://meritbbs.unimaas.nl/rmpdf/rm94_017.pdf, (2000): 1-39.

Voss, D. "'New Physics' Finds a Haven at the Patent Office." *Science* 284(May 1999): 1252-1254.

Waterson, M. "The Economics of Product Patents." *The American Economic Review* 80(4, September 1990): 860-869.

World Intellectual Property Organization (WIPO). *General Information*. Available on line at http://www.wipo.org/about-wipo/en/, (2001a).

World Intellectual Property Organization (WIPO). *Intellectual Property Protection Treaties: Patent Law Treaty*. Available on line at http://www.wipo.org/treaties/ip/plt/index.html, (2001b).

World Intellectual Property Organization (WIPO). *Intellectual Property Protection Treaties*. Available on line at http://www.wipo.org/treaties/ip/plt/, (2001c).

APPENDIX

- *The curvature of the function* $h(x)=\frac{r+\lambda(x,b)}{\lambda_x}-r\Pi_d$ *is negative,*

$h_{xx}=-\frac{\lambda_{xxx}(r+\lambda)+\lambda_x\lambda_{xx}}{(\lambda_x)^2}+\frac{2(\lambda_{xx})^2(r+\lambda)}{(\lambda_x)^3}\leq 0$, *for both the additive and the multiplicative formulations of the instantaneous probability of success,* λ.

Proof:

To prove the above statement the additively separable function $f_1:\lambda=x^\theta+\frac{1}{b}$ and the multiplicatively separable function $f_2:\lambda=\frac{x^\theta}{b}$ are used. Both functions satisfy all the theoretical assumptions concerning the instantaneous probability of success λ.

- λ is additively separable in x and b, $f_1:\lambda=x^\theta+\frac{1}{b}$.

In this case,

$$h_{xx}=(1-\theta)\{\frac{1}{x}-x^{-(1+\theta)}(\frac{1}{b}+r+x^\theta)\}=(1-\theta)\{\frac{1}{x}-(\frac{1}{bx^{(1+\theta)}}+\frac{r}{x^{(1+\theta)}}+\frac{1}{x})\}=$$
$$(1-\theta)(-\frac{1+rb}{bx^{(1+\theta)}})\leq 0$$

The above inequality holds $\forall$ $\theta\in(0,1)$, $b\in(0,1]$ and $r\in[0,1]$.

- λ is multiplicatively separable in x and b, $f_2:\lambda=\frac{x^\theta}{b}$.

In this case,

$$h_{xx}=(1-\theta)\{\frac{1}{x}-x^{-(1+\theta)}b(r+\frac{x^\theta}{b})\}=(1-\theta)\{\frac{1}{x}-(\frac{rb}{x^{(1+\theta)}}+\frac{1}{x})\}=(1-\theta)(-\frac{rb}{x^{(1+\theta)}})\leq 0$$

The above inequality holds $\forall$ $\theta\in(0,1)$, $b\in(0,1]$ and $r\in[0,1]$. □

Proposition 5.9 *An increase in the monopoly profits leads to an increase in the optimal patent breadth* ($\frac{db^*}{d\Pi_m} > 0$)*, when* $b^* \in (0, \bar{b})$ *and a decrease in the optimal patent breadth* ($\frac{db^*}{d\Pi_m} < 0$)*, when* $b^* \in (\bar{b}, 1]$. *The patent breadth* $\bar{b} \in (0,1]$ *is the breadth of patent protection that makes the effect of a change in monopoly profits on the optimal patent breadth equal to zero,* $\frac{d(b^* = \bar{b})}{d\Pi_m} = 0$. *The patent breadth* $\bar{b}$ *exists for both the additive and multiplicative formulations of the instantaneous probability of success.*

Proof:

To prove the above proposition the additively and multiplicatively separable functions $f_1 : \lambda = x^\theta + \frac{1}{b}$, $f_2 : \lambda = \frac{x^\theta}{b}$ are used, respectively. When the function f_1 is used $f_{\Pi_m} = -\frac{(b^2-1)}{(1+b(r+x^\theta))^2} - \frac{2b^2}{1+b(r+x^\theta)}$. The patent breadth $\bar{b}$ that makes $f_{\Pi_m} = 0$ is given by $\bar{b} = -\frac{1}{2(r+x^\theta)} + \frac{3}{2^{2/3}(r+x^\theta)A} + \frac{A}{62^{1/3}(r+x^\theta)}$ where, $A = \left(-54 + 108(r+x^\theta)^2 + \sqrt{-2916 + (-54 + 108(r+x^\theta)^2)^2}\right)^{1/3}$. Performing simulations it is found that there are combinations of $r \in [0,1]$, $x \geq 0$ and $\theta \in (0,1)$ values such that there exists a patent breadth $\bar{b} \in (0,1]$. The term f_{Π_m} is decreasing in patent breadth, $\frac{\partial f_{\Pi_m}}{\partial b} = -\frac{2\left(3b + r + x^\theta + 3b^2(r+x^\theta) + b^3(r+x^\theta)^2\right)}{\left(1+b(r+x^\theta)\right)^3} < 0$, $\forall$ $b \in (0,1]$, $\theta \in (0,1)$, $x \geq 0$ and $r \in [0,1]$. The above imply that if patent breadth b^* is such that $b^* \in (0, \bar{b})$ then $f_{\Pi_m} > 0$

which implies that $\frac{db^*}{d\Pi_m} > 0$ while if patent breadth b* is such that $b^* \in (\bar{b},1]$ then $f_{\Pi_m} < 0$ which implies that $\frac{db^*}{d\Pi_m} < 0$. When the multiplicatively separable function f_2 is used, $f_{\Pi_m} = -\frac{(b^2-1)x^\theta}{(br+x^\theta)^2} - \frac{2b^2}{br+x^\theta}$. The patent breadth $\bar{b}$ that makes $f_{\Pi_m} = 0$ is given by $\bar{b} = -\frac{x^\theta}{2r} + \frac{x^{2\theta}}{2rB} + \frac{B}{2r}$ where $B = \left(2r^2x^\theta - x^{3\theta} + 2\sqrt{r^4x^{2\theta} - r^2x^{4\theta}}\right)^{1/3}$. Performing simulations it is found that for certain r∈[0,1], x≥0 and θ∈(0,1) values $\bar{b} \in (0,1]$. The term f_{Π_m} is decreasing in patent breadth, $\frac{\partial f_{\Pi_m}}{\partial b} = -\frac{2\left(b^3r^2 + rx^\theta + 3b^2rx^\theta + 3bx^{2\theta}\right)}{\left(br+x^\theta\right)^3} < 0$, ∀ b∈(0,1], θ∈(0,1), x≥0 and r∈[0,1]. The above imply that if patent breadth b* is such that $b^* \in (0,\bar{b})$ then $f_{\Pi_m} > 0$ which implies that $\frac{db^*}{d\Pi_m} > 0$ and if patent breadth b* is such that $b^* \in (\bar{b},1]$ then $f_{\Pi_m} < 0$ which implies that $\frac{db^*}{d\Pi_m} < 0$. □

Proposition 5.10 *An increase in the duopoly profits leads to a decrease in the optimal patent breadth (*$\frac{db^*}{d\Pi_d} < 0$*), when* $b^* \in (0,\bar{\bar{b}})$ *and to an increase in the optimal patent breadth (*$\frac{db^*}{d\Pi_d} > 0$*), when* $b^* \in (\bar{\bar{b}},1]$. *The patent breadth* $\bar{\bar{b}} \in (0,1]$ *is the breadth of patent protection that makes the effect of a change in duopoly profits on the optimal* *patent breadth equal to zero,* $\frac{d(b^*=\overline{\overline{b}})}{d\Pi_d}=0$. *The patent breadth* $\overline{\overline{b}}$ *exists for both the additive and multiplicative formulations of the instantaneous probability of success.*

Proof:

When $f_1:\lambda=x^\theta+\frac{1}{b}$ is used $f_{\Pi_d}=\frac{2b}{r}+\frac{r(-1+b^2)}{(1+b(r+x^\theta))^2}-\frac{2b(1+bx^\theta)}{1+b(r+x^\theta)}$. The patent breadth $\overline{\overline{b}}$ that makes $f_{\Pi_d}=0$ is given by $\overline{\overline{b}}=-\frac{C}{6D}-\frac{\left(-(C^2+12(-1+r)D\right)}{E^{1/3}}+\frac{E^{1/3}}{62^{1/3}D}$

where, $C=\left(-4r+r^2-4x^\theta+4rx^\theta\right), D=-r^2-2rx^\theta+r^2x^\theta-x^{2\theta}+rx^{2\theta}$,

$E=32^{2/3}D\left((-2C^3)+36(-1+r)CD-108r^2D^2+F\right)$and

$F=\sqrt{4\left((-C)^2+12(-1+r)D\right)^3+\left(-2C^3+36(-1+r)CD-108r^2D^2\right)^2}$. Performing simulations it is found that for certain θ∈(0,1), x≥0 and r∈[0,1] values $\overline{\overline{b}}\in(0,1]$. The term f_{Π_d} is increasing in patent breadth ∀ b∈(0,1], θ∈(0,1), x≥0 and r∈[0,1],

$$\frac{\partial f_{\Pi_d}}{\partial b}=\frac{2\begin{pmatrix}(1+bx^\theta)^3-r(1+bx^\theta)^2(1+b(-3+x^\theta))+r^3(1-b^3(-1+x^\theta))+\\ r^2(x^\theta-3b^2(-1+x^\theta)-b^3x^9(-3+2x^\theta))\end{pmatrix}}{r\left(1+b(r+x^\theta)\right)^3}>0.$$

The above imply that if patent breadth b* is such that $b^*\in(0,\overline{\overline{b}})$ then $f_{\Pi_d}<0$ which implies that $\frac{db^*}{d\Pi_d}<0$ and for if patent breadth b* such that $b^*\in(\overline{\overline{b}},1]$ then $f_{\Pi_d}>0$ which implies that $\frac{db^*}{d\Pi_d}>0$. When the function $f_2:\lambda=\frac{x^\theta}{b}$ is used, $f_{\Pi_d}=\frac{2b}{r}+\frac{(-1+b^2)rx^\theta}{(br+x^\theta)^2}-\frac{2bx^\theta}{br+x^\theta}$. The patent breadth $\overline{\overline{b}}$ that makes $f_{\Pi_d}=0$ is given by $\bar{\bar{b}} = -\frac{(-4+r)x^{\theta}}{6r} - \frac{G}{32^{2/3}r^{2}(H+\sqrt{4G^{3}+H^{2}})^{1/3}} + \frac{(H+\sqrt{4G^{3}+H^{2}})^{1/3}}{62^{1/3}r^{2}}$ where

$G = (-(-4+r)^{2}r^{2}x^{2\theta} - 12(-1+r)r^{2}x^{2\theta})$ and

$H = (108r^{6}x^{\theta} + 2(-4+r)^{3}r^{3}x^{3\theta} + 36(-4+r)(-1+r)r^{3}x^{3\theta})$. Performing simulations it is found that for certain $\theta \in (0,1)$, $x \geq 0$ and $r \in [0,1]$ values $\bar{\bar{b}} \in (0,1]$. The term f_{Π_d} is increasing in patent breadth, $\frac{\partial f_{\Pi_d}}{\partial b} = \frac{2b^{3}r^{3} + 3b^{2}r^{2}x^{\theta} + r^{3}x^{\theta} + 3brx^{2\theta} + x^{3\theta} - rx^{3\theta}}{r(br+x^{\theta})^{3}} > 0$, $\forall\, b \in (0,1]$, $\theta \in (0,1)$, $x \geq 0$ and $r \in [0,1]$. The above imply that if patent breadth b^{*} is such that $b^{*} \in (0,\bar{\bar{b}})$ then $f_{\Pi_d} < 0$ which implies that $\frac{db^{*}}{d\Pi_d} < 0$ and for if patent breadth b^{*} such that $b^{*} \in (\bar{\bar{b}},1]$ then $f_{\Pi_d} > 0$ which implies that $\frac{db^{*}}{d\Pi_d} > 0$. □